Curtin University

Research article

Parameter estimation of industrial PET reactor using multi-objective kernel density estimation of distribution algorithm

Na Luo[1,2] *and Feng Qian*[1,2]*

[1]*Key Laboratory of Advanced Control and Optimization for Chemical Processes, East China University of Science and Technology, Ministry of Education, Shanghai, 200237, China*
[2]*State-Key Laboratory of Chemical Engineering, East China University of Science and Technology, Shanghai, 200237, China*

ABSTRACT: With a multitude of reaction pathways, poly (ethylene-terephthalate) (PET) polymerization of industrial practice is complex, and the quality of PET is normally described in terms of several experimentally measured indices. In this paper, parameters estimation of industrial PET reactors is presented as a multi-objective problem to make the mathematic model consistent with the actual industrial process. Considering the interrelation among parameters and the failure of general optimization algorithms, a new multi-objective estimation of distribution algorithm is proposed. Kernel density estimation is used to make the new population more suitable for real-life problems instead of Gaussian model during the evolution of the algorithm. Strategies including selection of kernel width, sampling method and Pareto domination selection are used to explore and exploit the search space more efficiently. With industrial operating data identified in steady state and eliminated from gross error, kinetic parameters are estimated by minimizing carboxyl end group concentration and degree of polymerization simultaneously using the proposed algorithm. The simulation results show that the model with estimated parameters has better predictive performance compared with the experimental parameters. Copyright © 2011 Curtin University of Technology and John Wiley & Sons, Ltd.

Keywords: poly(ethylene terephthalate); estimation of distribution algorithm; parameter estimation; Pareto solution

INTRODUCTION

Mathematical models are important in gaining insight into the mechanism of chemical process. When they are used for industrial processes, it is necessary to validate the model from historical plant data using estimation method for kinetic parameters. In general, the parameter estimation of a model is based on the minimization of some quantities. For linear models, the parameter estimation is easy when using linear regression or plot procedures; whereas for nonlinear models, it is difficult to choose an effective approach. Recently, many evolutionary algorithms such as genetic algorithm, [1] particle swarm optimization, [2] DNA computing, [3] differential evolution algorithm [4] and simulated annealing algorithm [5] were investigated aiming at the verification of the parameters and global optimization of the solution space.

As an important raw material for the production of synthetic fibers, films and beverage bottles, process of poly (ethylene-terephthalate) (PET) is very complex in industrial practice. Hence, the study of PET has attracted the attention of many. In the aspect of modeling, Yamada *et al*. [6] developed a simulation model for the three-phase continuous esterification process of terephthalate (TPA) with ethylene glycol (EG) that could express the oligomer properties and the concentration of each composition in the vapor and liquid phase. Ravindranath and Mashelkar [7] also proposed the mathematical model of a continuous polycondensation reactor, and the predictions have been compared with available pilot plant data. At the same time period, Gupta [8] had some research on the modeling, which was tested on some industrial scale reactors. Bhaskar *et al*. [9] and Ramteke *et al*.[10] continued the work with improving the model. Kang *et al*.[11] presented a comprehensive kinetic model for the direct esterification reaction of PET synthesis in semibatch and continuous reactors using the functional group analysis. Taking into account the presence of evaporable components in the reaction mixture, Dorogov [12] provided a mathematical model for the monomerization stage of PET production. In the model developed by Patel *et al*.[13],

Correspondence to: Qian Feng, 130 Meilong Road, Shanghai, 200237, China. E-mail: fqian@ecust.edu.cn

rate constants were optimized by data fitting with the oligomeric chain length. Although there are so many researches on PET modeling, the approaches can be classed into two types: molecular species models and functional group models. In the first approach, all molecular species are considered, whereas in the second approach, only functional groups are treated as special entities.

For industrial PET process, the modeling validation problem has been scaled because of the fact that most researchers do not have access to industrial data sets. Parameters are determined by using the laboratory experimental data, and they expand to simulate industrial process. In modeling for practical process, thermodynamics and the kinetics of reactions are limited by many factors such as the mixing condition of the TPA and EG, the solubility of TPA in the mixture, so that even if clear data is available in literature, the parameters are needed to be corrected using the industrial operating data.

Though parameter estimation of PET reaction scheme is rarely discussed, there is much research about optimization of polymer process, which can give us some suggestions. Bhaskar et al.[14,15] applied the adapted non-dominated set-ring genetic algorithm into the multi-objective optimization of wiped-film PET reactor. The acid and vinyl end-group concentrations in the product were both selected as the objectives. Multi-objective differential evolution algorithm [16] was applied to solve the multi-objective optimization problem of wet film PET reactor considering minimization of acid end group and vinyl end group as the main objectives. In the integrated multi-level optimization of PET plant, [17] the objective function was the integration of several objectives. On the whole, just as Bhaskar et al.[18] pointed, the optimization of PET process should consider several properties simultaneously. Hence, parameter estimation of PET should be a multi-objective problem.

Since PET process is high nonlinear, and the parameters are strong correlative, traditional optimization techniques often fail to provide the best (global) solution. To overcome these difficulties, heuristic optimization methods, such as genetic algorithm (GA), particle swarm optimization algorithm (PSO), simulated annealing algorithm (SA) or differential evolutionary algorithm (DE) are considered. For real industrial optimization problems, especially in chemical process, these standard algorithms should be improved using self-adaptive strategy or combined with other evolutionary algorithms to obtain satisfactory results. Unlike derivative-based and direct search methods, these algorithms do not require initial guesses for model parameters and do not use derivatives. [2] They assure a higher probability to find the global minima by the large number of function evaluations and a random search character. With parallel computation of population, Pareto set can be efficiently obtained using the multi-objective algorithms. As a new developed evolutionary algorithm, estimation of distribution algorithms (EDAs) has drawn significant attention for global optimization. Compared with GA, PSO and SA, EDA allows for less parameters setting with better performance. Consequently, only EDAs will be discussed in detail in this work.

The core of EDA is the probabilistic model. In continuous EDAs, Gaussian model is the predominantly used probabilistic model. However, the assumption of normality is not realistic for many real-life problems. In this paper, a new multi-objective estimation of distribution algorithm named MKEDA is proposed. The proposed algorithm is based on kernel density estimation models. Strategies including selection of kernel width, sampling method and Pareto domination selection are included in the proposed algorithm to efficiently explore and exploit the search space. The performance of the MKEDA is tested through several benchmark functions. Accordingly, it can be used in parameter estimation of nonlinear, nonconvex and complex PET reactor model.

This paper is organized as follows. In section 2, EDA is reviewed. Section 3 elaborates on MKEDA and tests its performance on several benchmark functions. Parameters estimation problems in PET reaction scheme are proposed and analyzed in section 4. In the end, the conclusion is drawn.

ESTIMATION OF DISTRIBUTION ALGORITHM

Estimation of distribution algorithms (EDAs) is a class of probabilistic and graphical model-based evolutionary computational methods, which receive increasing attention recently. Different from traditional genetic algorithms, there is neither crossover nor mutation operator in EDAs, and new candidate individuals at the next generation are reproduced through sampling from a probability distribution of promising individuals. A general outline of EDA is as follows:

(1) Initialize a population of individuals randomly.
(2) Select individuals for modeling and elite individuals.
(3) Estimate parameters of probabilistic model using the selection set.
(4) Sample new individuals from probabilistic model as new generation.
(5) New generation is partially replaced by the elite individuals.
(6) Stop if some stopping criterion is reached, else go to step 2.

Probabilistic model is the core of EDAs. In the combinatory domain, EDAs are categorized into three classes according to the interaction between variables. In the first class, variables are assumed independent, and the probabilistic model is developed using a product of independent univariate probabilities. Probabilistic

incremental learning (PBIL),[19] compact genetic algorithm (CGA)[20] and univariate marginal distribution algorithm (UMDA)[21] are included in this class. The second class disposes variables as bivariate dependence using Bayesian deduction, which includes mutual information maximization for input clustering (MIMIC),[22] combining optimizers with mutual information trees algorithm (COMIT)[23] and bivariate marginal distribution algorithm (BMDA).[24] Extended compact genetic algorithm (ECGA),[25] the Bayesian optimization algorithm (BOA)[26] and the factorized distribution algorithm (FDA)[27] are the third class, which build the probability model using multivariate dependent Bayesian network. Based on these combinatory optimization algorithms, UMDA$_C$,[28] PBIL$_C$,[29] MIMIC$_C$,[28] EGNAee[28] are extension EDAs of UMDA, PBIC, MIMIC, and EGNA in continuous domain based on Gaussian network.

In practice, the interactions of variables are often unclear in advance for real problems. At the same time, complex learning algorithms will lead to significant additional computation, but not better performance. So, it is important to construct suitable probabilistic models for real-world problems.

MULTI-OBJECTIVE KERNEL-DENSITY ESTIMATION OF DISTRIBUTION ALGORITHM (MKEDA)

Multi-objective optimization problem

Different from single-objective problems, multi-objective problems simultaneously optimize multiple objectives, which are conflicting. Without loss of generality, assuming that there are D objectives $f_i(x)$, $i = 1, \cdots, D$ to be minimized, the problem is defined as follows:

$$
\begin{aligned}
&y_i = f_i(x), \qquad i = 1, \ldots, D \\
&\text{Minimize } y = f(x) = (f_1(x), \ldots, f_D(x)) \\
&\text{subject to } e(x) = (e_1(x), \ldots, e_J(x)) \geq 0 \\
&x \in S
\end{aligned} \tag{1}
$$

where D is the number of the objective functions, $f_i : \mathbb{R}^p \to \mathbb{R}. x = [x_1, \cdots, x_p]^T$ is the vector of decision variables with the feasible domain. The particular set of values $x_1^*, \cdots, x_p^*$ should be determined to yield the optimum values of all the objective functions from the set S of all vectors, which satisfy J constraints $e_j(x) \geq 0$, $j = 1, \ldots, J$.

Since the objectives are competing, it is a challenge for optimization techniques to scalably discover an optimal trade-off between objectives. There are two basic approaches to solving multi-objective optimization problems. The first is to yield a single-objective problem by weighing the objectives. The other

approach is to find the Pareto-optimal front, which is used as the most commonly accepted term of optimality. A vector of decision variables $x* \in S$ is Pareto optimal, if there does not exist another $x \in S$ such that $f_i(x) \leq f_i(x*)$ for all $i = 1, \ldots, D$ and $f_j(x) < f_j(x*)$ for at least one j. The Pareto optimal is not a single solution, but a set of solutions.

Non-dominated sorting genetic algorithm-II (NSGA-II)[30] is the most popular algorithm, which is the extension of genetic algorithm to multi-objective domain. For other heuristic algorithms, they are also extended to multi-objective domains such as multi-objective PSO (MPSO).[31] Based on the framework of probability model, Bosman[32] proposed multi-objective mixture-based iterated density estimation evolutionary algorithm, which a specialized diversity preserving selection operator. Laumanns[33] incorporated the selection and replacement operators of strength Pareto evolutionary algorithm-2 (SPEA2) into mixed BOA, which extends BOA with decision graphs to solve problems with both discrete as well as continuous variables. In multi-objective parzen-based estimation of distribution algorithm, Costa[34] used the ranking method of NSGA-II and the Parzen estimator to approximate the probability density of solutions lying on the Pareto front. Based on PBIL, Bureerat[35] presented the corresponding algorithm for multi-objectives. Also, UMDA are also extended to MUMDA.[36] Although there are so many kinds of multi-objective optimization algorithms, they should be adapted for new problems. That is why a new algorithm is presented in this paper.

Multi-objective kernel-density estimation of distribution algorithm

Gaussian model is simple and easy to understand in continuous domain, whereas the assumption of normality is not realistic for many real-life problems. Hybrid Gaussian model was proposed to compensate shortage of Gaussian model.[37]. Non-parametric histogram model has also been used in EDA without any assumption of population distribution.[38,39] However, these models scale up exponentially with increment of problem size. Compared with these probabilistic models, kernel density estimation has more interesting properties such as intrinsic multimodality, which are proper to describe the solution distribution of complex and multimodal continuous problems with lower computational complexity. Bosman[40] proposed an iterative density estimation of evolutionary algorithm, which used the kernel density estimation and extended it to multi-objective algorithm.[32] Although the idea is similar with MKEDA, the strategies such as kernel bandwidth determination, sampling methods and

Pareto domination selection is different in MKEDA. The framework of MKEDA is as follows:

(1) Initialize a population of individuals randomly.
(2) Calculate the fitness of population, select individuals D_s according to Pareto domination selection method.
(3) Select elitism individuals D_e from the generation using Pareto domination selection method.
(4) Compute kernel bandwidth according to D_s and the previous kernel bandwidth.
(5) Estimate kernel density model $\hat{p}(x) = \frac{1}{nh_n}\sum_{i=1}^{n}k\left(\frac{x-x_i}{h_n}\right)$.
(6) Sampling individuals from $\hat{p}(x)$ using stochastic universal sampling.
(7) Replace worst individuals using D_e.
(8) Stop if some stopping criterion is reached; otherwise go to (2).

In MKEDA, the strategies such as kernel bandwidth selection method, sampling method and Pareto domination selection are as follows:

Selection method of kernel bandwidth

In statistics, kernel density estimation is a non-parametric way to estimating the probability density function of a random variable. The kernel density estimation of distribution $\hat{p}(x)$ is defined as:

$$\hat{p}(x) = \frac{1}{nh_n}\sum_{i=1}^{n}k\left(\frac{x-x_i}{h_n}\right) \qquad (2)$$

where $k(u)$ is a kernel function, h_n is kernel width which is a smooth parameter related with the samples n. The kernel function has a bounded $k(u) \geq 0$ and satisfies the following conditions:

$$(1)\int_{-\infty}^{+\infty}|k(u)|\mathrm{d}u < +\infty \quad (2)\lim_{|u|\to\infty}uk(u) = 0 \quad (3)k(-u)$$
$$= k(u) \quad (4)\int_{-\infty}^{+\infty}k(u)\mathrm{d}u = 1$$

Kernel functions can be uniform, triangle, epanechnikov, quadratic, tri-weight, Gaussian, cosines or exponent function. Among these functions, Gaussian kernel is the most commonly used:

$$k(x, x_i) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-x_i)^2} \qquad (3)$$

In kernel density estimation, the key factor is not the type of kernel function, but the kernel bandwidth. Figure 1 illustrated different estimated distribution model for the same samples. It is clear that the
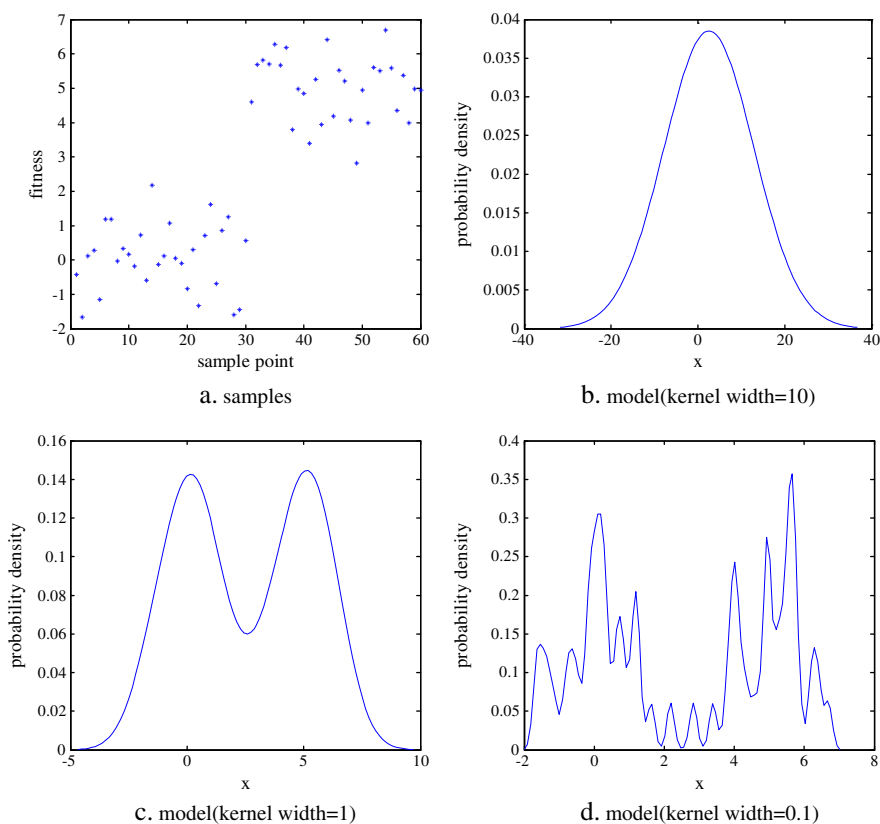

Figure 1. Distribution model of samples. This figure is available in colour online at www.apjChemEng.com.

bandwidth of the kernel intensively influences the shape of distribution function. If the bandwidth is too narrow, the density estimate will have a choppy appearance, and if the bandwidth is too wide, the estimate will be overly smooth and interesting features of the density are obscure.

Gaussian kernel is used as kernel function in this paper. As for the selection of bandwidth, a rule of thumb proposed by Hardle [41] is used for single-variable:

$$h_n = \frac{1.06}{n^{1/5}}\sigma \qquad (4)$$

where σ is the sample standard deviation, n is the number of samples. The kernel bandwidth is based on the optimal criterion, which minimizes integral of mean square error. For static data, the rule of bandwidth selection is valid. However, it is not enough for dynamic evolutionary process. In EDAs, the initial probability distribution is the prior distribution, whereas the next generation is the posterior distribution which is calculated using the prior distribution. It is obvious that the EDA, which only uses the probability model of the current population generates the next generation that is not enough. Therefore, the kernel bandwidth in the probability model should be updated according to two kinds of information: historical and current information. The kernel bandwidth is used as follows:

$$h_n(j) = \alpha h_n(j-1) + (1-\alpha)\hat{h}_n(j) \qquad (5)$$

where $h_n(j)$ is the kernel width in the current model, $h_n(j-1)$ is the kernel width in the history model, and $\hat{h}_n(j)$ is the estimated kernel width by Equation (4). $\alpha(0 \le \alpha \le 1)$ is the accumulation factor, which determines the effect of the history model on the current model. j is the number of the generation. Through accumulation of the history kernel bandwidth, the information of the historical model and the good individuals in the past generations is reserved.

Sampling method

Sampling method is also important in MKEDA. Reducing sampling error is the standards of choosing sampling method. We used extended stochastic universal sampling method (E-SUS) [38] as our sampling method. E-SUS extended Baker's stochastic universal sampling method was proposed for the proportional selection operator. When used in this paper, the kernel density function is necessary to be divided into many bins. For each new individual, each value is generated as follows: first, a bin is selected according to the probability of each bin. Then, the value of variable is generated by generating a number from the selected bin with uniform distribution. This is repeated until all individuals are obtained.

Pareto domination selection and diversity preservation

To identify solutions of the Pareto front in the population, each solution should be compared with every other solution in the population to find out if it is dominated. The computational complexity is high. Pareto domination tournament is used in this paper to select the Pareto front. Two candidates are selected from the current generation randomly, and a comparison set of individuals are picked randomly from the elitism set. The two candidates are compared against each other. The non-dominated candidate is then compared against each individual in the comparison set. If the candidate is not dominated by the comparison set, it is selected into the elitism set, and vice versa. To circumstance the elitism set in a fixed set and preserve the diversity, niche fitness sharing is used to select fixed elitism set. The sharing function method involves a sharing parameter, which sets the extent of sharing desired in a problem. This parameter is related to the distance metric chosen to calculate the proximity measure between two population members. The parameter denotes the largest value of that distance metric within which any two solutions share each other's fitness. It is defined as follows:

$$F(\mathrm{Pop}(i)) = \frac{1}{\displaystyle\sum_{\mathrm{Pop}(j)\in P} s(d(\mathrm{Pop}(i),\mathrm{Pop}(j)))} \qquad (6)$$

where P is the set of elitism, $s(d(\mathrm{Pop}(i),\mathrm{Pop}(j)))$ is the sharing function of individuals $\mathrm{Pop}(i)$ and $\mathrm{Pop}(j)$.

$$s(d(\mathrm{Pop}(i),\mathrm{Pop}(j))) = \begin{cases} 1 - \left(\dfrac{d(\mathrm{Pop}(i),\mathrm{Pop}(j))}{\sigma_{\mathrm{share}}}\right)^{\alpha}, & \text{if } d(\mathrm{Pop}(i),\mathrm{Pop}(j)) < \sigma_{\mathrm{share}} \\ 0, & \text{else} \end{cases} \qquad (7)$$

where $d(\text{Pop}(i), \text{Pop}(j))$ is the distance between individuals $\text{Pop}(i)$ and $\text{Pop}(j)$. It is calculated as Euclidean distance and σ_{share} is niche radius, fixed by average distance of the individuals in population. Individuals within σ_{share} distance of each other degrade each other's fitness, because they are in the same niche. Thus, convergence occurs within a niche, but convergence of the full population is avoided. As one niche 'fills up',, its niche count increases to the point that its shared fitness is lower than that of other niches. [42]

Performance of MKEDA

To test the proposed algorithm MKEDA and compare it with related works, five benchmark functions are adopted, namely Schaffer, FON, KUR, ZDT4 and ZDT6. These functions include characteristics that are suitable for examining the effectiveness of multi-objective approaches in maintaining the population diversity as well as converging to the Pareto front. All problems have two objective functions, and none of them has any constraint. The performance of MKEDA are compared with NSGA-II, [30] MPSO, [31] MUMDA. [36] In all algorithms, the population size is set to 100, and each run continues until a maximum of 10,000 function evaluations is reached. The initial population is generated uniformly. Other specific parameters are listed in Table 1.

Ability to find Pareto solutions

The experimental results of the five benchmark functions are shown in Figure 2. The subfigure is the result of NSGA-II, MPSO, MUMDA and MKEDA from left to right. For Schaffer problem, all algorithms can converge to the Pareto front, whereas the results of NSGA-II are well distributed. For FON problem, all algorithms can find solutions close to the Pareto front except MPSO. The MKEDA show better performance in the KUR, ZDT4 and ZDT6 test functions comparing with other algorithms. From all the test performances, it showed that MKEDA can find solutions, which closely

Table 1. Parameters setting.

	Selection parameter: 0.3
	Crossover operator: 5
NSGA-II	Mutation operator: 5
MPSO	Inertial weight: 0.729 Accelerate coefficient 1: 1.495 Accelerate coefficient 2: 1.495
MUMDA	Truncation ratio for selection: 0.3 Elitism population: 5
MKEDA	Truncation ratio for selection: 0.3 Elitism population: 5

NSGA-II, Non-dominated sorting genetic algorithm-II; MPSO, multi-objective particle swarm optimization; MUMDA, multi-objective univariate marginal distribution algorithm; MKEDA, multi-objective estimation of distribution algorithm.
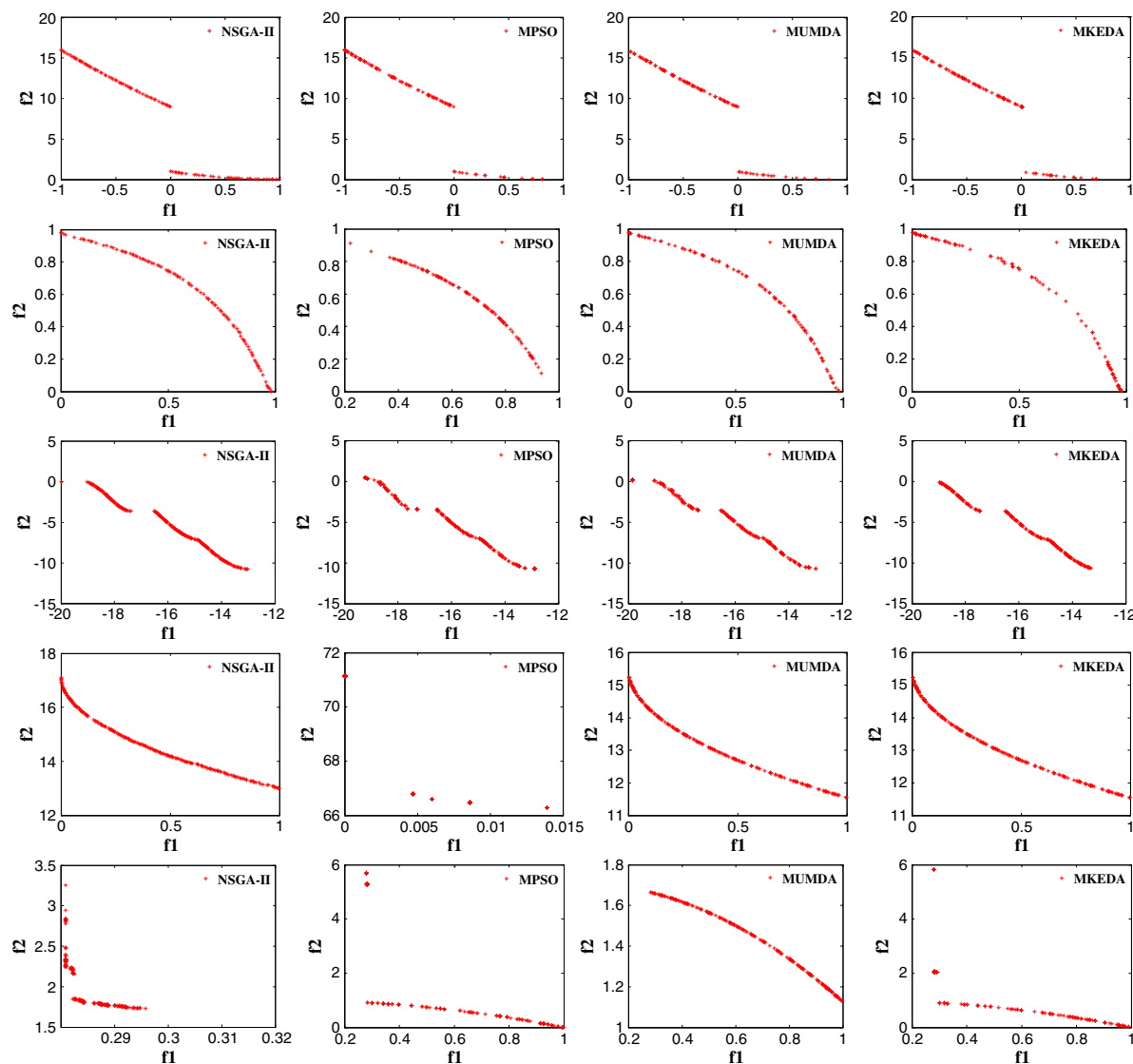
resemble the Pareto front solutions and the uniformity of the solutions are satisfactory.

Ability to adaptive clustering

It is well known that clustering in the evolution is beneficial to find global best solutions in the multi-modal problems. In MKEDA, the use of kernel density estimation provides the algorithm the characteristics of adaptive clustering. Considering the complexity of multi-objective benchmark functions, the Griewank multimodal function is used to demonstrate this ability. Figure 3 shows the distribution of individuals at the first, tenth, 20th and 1000th generation during the optimization of Griewank multimodal function.

In the beginning of the algorithm, the individuals scatter in the whole space randomly. While the algorithm continues, the individuals are clustered adaptively with several centers. Around the centers, the algorithm keeps on finding the local optimal. The connection between centers also makes the algorithm easy to escape from local optimal. When the algorithm finds the maximum probability of the global best solution, it will dig in the center and the surroundings until better solutions are found. So, MKEDA has more probability to escape from the local optimal and find the global solution.

PARAMETERS ESTIMATION OF INDUSTRIAL PET REACTOR

PET reaction scheme

PET is produced by the reactions of TPA and EG. The process including three stages: esterification, pre-polycondensation and polycondensation. Catalysts are added in the middle of the process, and the mass transfer should be considered in the finisher stage. In the whole process, reaction scheme can be assumed the same.

In this study, functional group approach is used to establish the reaction scheme. Polymerization is regarded as reactions between two reactive functional groups. All components considered in the reaction scheme are listed in Table 2. Different segments are defined in which 't' and 'b' represent 'terminal functional group' and 'bound monomer repeating units', respectively. In PET process, there are two kinds of main reactions: esterification and polycondensation accompanied by several side reactions shown in Table 3. Esterification reactions are those generating water as by-product, which include reactions 1–4 in Table 3, and reaction 5 is polycondensation reaction, which generate EG. Also, the last three reactions are the side reactions leading to DEG formation.

For the calculation of reaction rates, it is assumed that functional group reactivity does not depend on the polymer chain length. Hence, the material balance

Figure 2. Obtained non-dominated solutions on Schaffer, FON, KUR, ZDT4 and ZDT6 problems This figure is available in colour online at www.apjChemEng.com.

equations of each component are similar to the normal reactions, which are not discussed here. The reaction rates for reactions 1–8 are shown in Table 4. The rate constants follow the well-known Arrhenius equation $k_i = k_{i0}e^{-E_a/RT}$ in which k_{i0} is the pre-exponential factor, E_a is the activation energy, R is the idea gas constant, and T is the absolute temperature.

Physical properties of PET are calculated based on segments. Polymer non-random two-liquid (NRTL) activity coefficient models are used to compute the properties of polymers. These physical models extend NRTL model from low molecular weight compounds to segment approach based on non-random concept and differ primarily from Flory–Huggins model in which binary interaction parameters are dependent of the concentration and molecular weight of components.

Proposition of parameter estimation problem in PET process

Parameter estimation is carried out for PET reaction scheme to validate the process model. In the Arrhenius equation, the activation energies E_a have been fixed based on previous simulation and experimental studies. The pre-exponential factors k_{i0} of the kinetic parameters are estimated next from historical plant operating data. Because the generation of DEG is much smaller than EG, rate constants of reactions 6–8 maintain the same results as that of lab experiments. To narrow the range of the pre-exponential factors, we choose parameters published [17] as our reference and the intervals of estimated variables are expanded to ±130% of the published values, shown in Table 5.

a. iteration=1

b. iteration=10

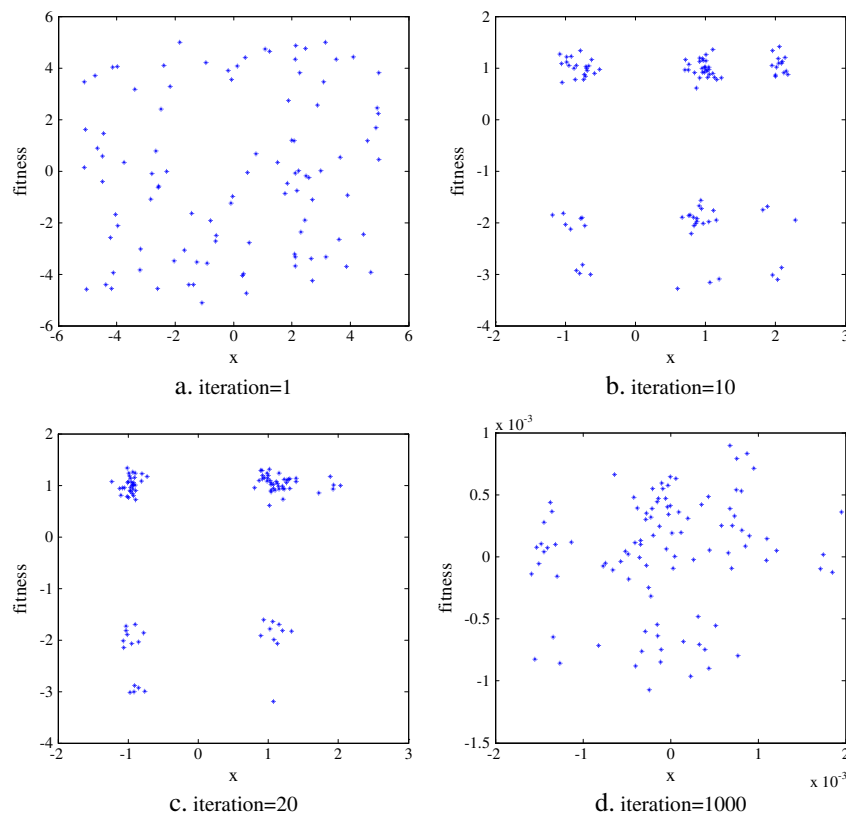c. iteration=20

d. iteration=1000

Figure 3. Distribution of new individuals using kernel density estimation. This figure is available in colour online at www.apjChemEng.com.

Table 2. Molecular structure of kinetic scheme components.

Acronym	Description	Molecular structure
EG	Ethylene glycol	$HOCH_2CH_2OH$
TPA	Terephthalic acid	$HOOC\text{-}\bigcirc\text{-}COOH$
tEG	EG-end group	$\sim OCH_2CH_2OH$
bEG	EG-repeat group	$\sim OCH_2CH_2O\text{-}R'$
tTPA	TPA-end group	$\sim OC\text{-}\bigcirc\text{-}COOH$
bTPA	TPA-repeat group	$\sim OC\text{-}\bigcirc\text{-}CO\sim$
H2O	Water	H_2O
DEG	Di-ethylene glycol	$HO(CH_2)_2O(CH_2)_2OH$
tDEG	DEG-end group	$\sim O(CH_2)_2O(CH_2)_2OH$
bDEG	DEG-repeat group	$\sim O(CH_2)_2O(CH_2)_2O\sim$

Considering the characteristics of PET process, carboxyl end group concentration $[C_{COOH}]$ and the degree of polymerization DP are simultaneously selected as objectives to control the esterification and polycondensation reaction rate. For each objective, the sum of squared errors between the online measurement and predicted values is minimized with respect to the pre-exponential factors. As for the online measurements, the errors are assumed to be independent and normally distributed. The multi-objective optimization problem is described mathematically as:

$$
\begin{aligned}
&Min\ I(k_{10}, k_{20}, k_{30}, k_{40}, k_{50}, k_{60}) \\
&= \min[Obj_1, Obj_2]^T \\
&= \min\left[\sum_{i=1}^{n} \left(\frac{[C_{COOH}]_{out,i} - [C_{COOH}]_{p,i}}{\overline{[C_{COOH}]_{out}}} \right)^2, \right. \\
&\quad \left. \sum_{i=1}^{n} \left(\frac{DP_{out,i} - DP_{p,i}}{\overline{DP_{out}}} \right)^2 \right]^T
\end{aligned}
\tag{8}
$$

Table 3. Reaction scheme and reaction constants.

| No. | Reactions | Rate constants | |
		Forward	Reverse
1	$tTPA + EG \rightleftharpoons bTPA + tEG + H_2O$	k_1	k_2
2	$tTPA + DEG \rightleftharpoons$ $bTPA + tDEG + H_2O$	k_1	k_2
3	$tTPA + tEG \rightleftharpoons bTPA + bEG + H_2O$	k_3	k_4
4	$tTPA + tDEG \rightleftharpoons$ $bTPA + bDEG + H_2O$	k_3	k_4
5	$tEG + tEG \rightleftharpoons bEG + EG$	k_5	k_6
6	$EG + EG \rightarrow DEG + H_2O$	k_7	—
7	$EG + tEG \rightarrow tDEG + H_2O$	k_8	—
8	$tEG + tEG \rightarrow bDEG + H_2O$	k_9	—

Table 4. Reaction rates.

Reaction rates

$R_1 = 2k_1[EG][tTPA] - k_2[tEG][H_2O]$
$R_2 = 2k_1[DEG][tTPA] - k_2[tDEG][H_2O]$
$R_3 = k_3[tEG][tTPA] - 2k_4[bEG][H_2O]$
$R_4 = k_3[tDEG][tTPA] - 2k_4[bDEG][H_2O]$
$R_5 = k_5[tEG][tEG] - 4k_6[bEG][EG]$
$R_6 = k_6[EG][EG]$
$R_7 = k_7[EG][tEG]$
$R_8 = k_8[tEG][tEG]$

Subject to:

$$[C_{DEG}]_{out} = [C_{DEG}]_d \qquad (9)$$

In the equations, subscripts 'out' and 'p' refer to the measurement values at the outlet of the reactor and the predicted values of PET property, respectively. $\overline{[C_{COOH}]_{out}}$ and $\overline{DP_{out}}$ are the average values of $[C_{COOH}]_{out}$ and DP_{out}. n is the number of the steady state used for parameters estimation. For PET process, the DEG generation is often considered. So, the concentration of DEG $[C_{DEG}]$ is incorporated in the objective functions as penalty functions. The new objectives are represented as:

$$\text{Min } Obj_1{}^* = \omega_1 \qquad (10)$$
$$\times \sum_{i=1}^{n} \left(\frac{[C_{COOH}]_{out,i} - [C_{COOH}]_{p,i}}{\overline{[C_{COOH}]_{out}}} \right)^2$$
$$+ \omega_3$$
$$\times \sum_{i=1}^{n} \left(\frac{[C_{DEG}]_{out,i} - [C_{DEG}]_{p,i}}{\overline{[C_{DEG}]_{out}}} \right)^2$$

$$\text{Min } Obj_2{}^* = \omega_2 \qquad (11)$$
$$\times \sum_{i=1}^{n} \left(\frac{DP_{out,i} - DP_{p,i}}{\overline{DP_{out}}} \right)^2$$
$$+ \omega_3$$
$$\times \sum_{i=1}^{n} \left(\frac{[C_{DEG}]_{out,i} - [C_{DEG}]_{p,i}}{\overline{[C_{DEG}]_{out}}} \right)^2$$

where $\omega_1, \omega_2, \omega_3$ are the penalty coefficients which are set as 0.7, 0.7, 0.3. $\overline{[C_{DEG}]_{out}}$ is the average value of $[C_{DEG}]_{out}$. With penalty method, the original problem is transferred to an unconstrained problem. This ensures that worse individual can be reproduced in the successive evolution of the algorithm and keeps the diversity of the population. Minimizations of $Obj_1{}^*$ and $Obj_2{}^*$ simultaneously lead to decrease the differences between the predictions and industrial operating data, whereas they give preference to be consistent with the real process.

Solutions and analysis

In this paper, esterification reactors are chosen to estimate the parameters of PET reaction scheme. Industrial operating data such as temperatures, pressures, flows, etc. are acquired every 5 min through distributed control system. In complex industrial environment, these dynamic data are often corrupted by random noise and systematic errors. Steady state identification (SSI) and gross error detection (GED) are needed to dispose industrial data before the parameters' estimation.

Various techniques are available for steady-state identification. [43] The most common and simple steady-state detectors analyze the data over a predefined moving window, as illustrated in Figure 4. In PET process, the residence time of material in a reactor is no longer than 60 min. So we predefined time interval 250 min. The moving window replaces each data point within the time interval 250 min and the moving window average is equivalent to a low-pass filter. The steady-state detector uses standard deviation $\hat{\sigma}$ within a steady state period and the difference Δ between the averages of two consecutive steady state periods as the criteria in a recursive fashion. If $\hat{\sigma}$ or Δ is larger than the minimal allowed values, the current data is in the unsteady state, and the identification turns to the next data.

Once the existence of the steady state is ascertained, the associated gross and random errors are further

Table 5. Intervals of estimated variables.

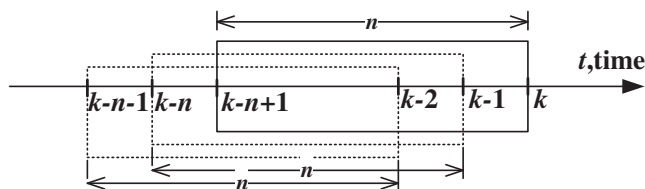| Rate constant sets	Intervals of k_{i0} ($\times 10^6$)	Ea(kcal/ kmol)
1	[1.456,2.704]	17.6
2	[1.456,2.704]	17.6
3	[0.5824,1.0816]	17.6
4	[1.1648,2.1632]	17.6
5	[0.952,1.768]	18.5
6	[1.904,3.536]	18.5

Figure 4. Moving windows of n data points at near kth time. This figure is available in colour online at www. apjChemEng.com.
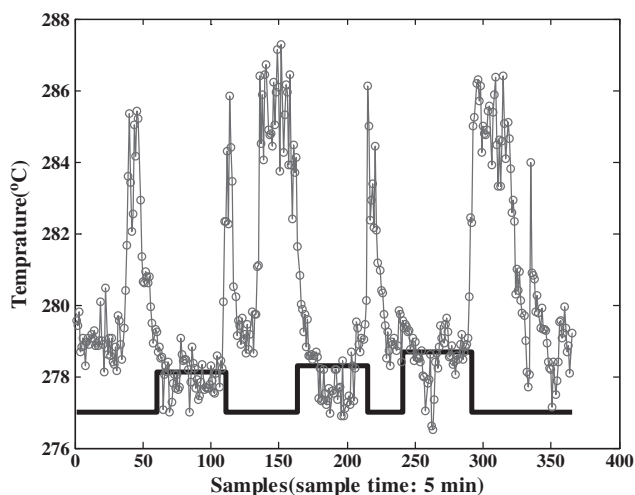

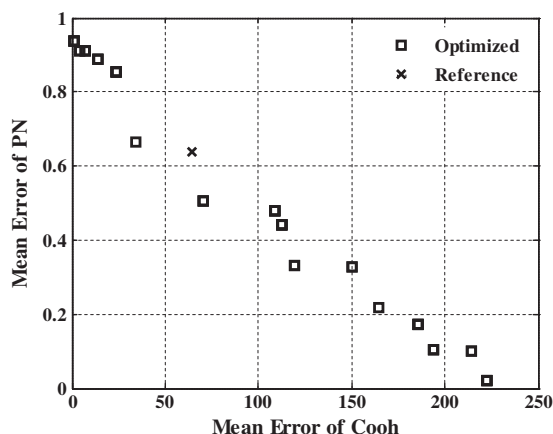**Figure 5.** Steady-state detection profile on reactor temperature.

Figure 6. Comparisons of objectives between estimation and original parameters in Yamada *et al*. [6]
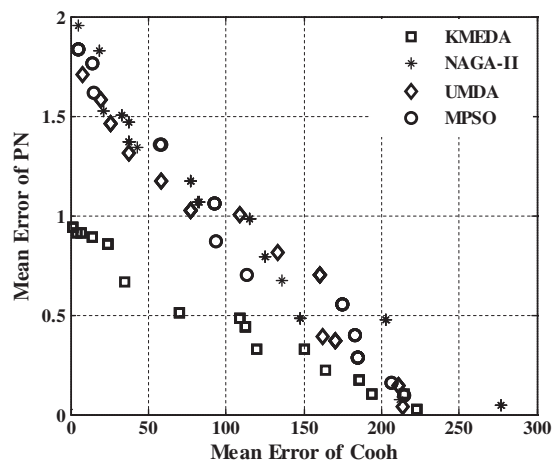

Figure 7. Comparisons of results between KMEDA, NSGA-II, UMDA and MPSO.

identified, and the magnitude of the gross error is taken as the difference between the measured and reference value. Data with gross errors are abandoned, and the average values are calculated for the steady state.

In PET process, reactor temperature data are the most sensitive factor, so they are used to identify the steady state of the whole process. With temperature data acquired from industrial process, the steady-state detection profile is shown in Figure 5 using the moving window and comparison strategy. The SSI flag indicates the state of the plant, and the value is the mean value without gross error. If the SSI flag is at the high level, it means the variable reached steady state and vice versa.

In the industrial process, variables of reactor pressure, temperature, residence time of the reaction mass inside the reactor, mass flow of PTA and EG are acquired for a period of time. Data reconciliation is used for the mass with the constraint of mass balance.
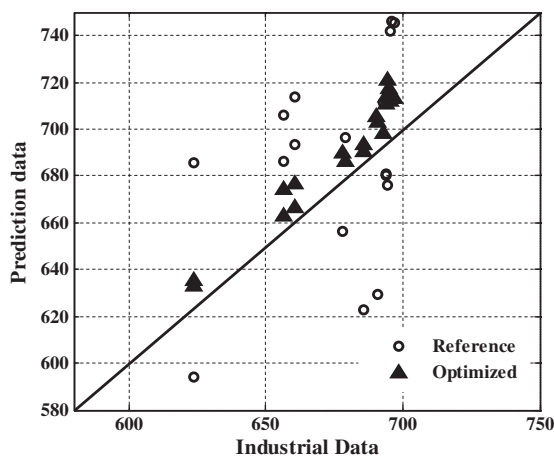
Figure 8. Parity plot for predictions with the optimized parameters and the data in reference.

removed using GED method. After industrial data are sorted, the minimum and maximum values are tested whether they are within the three-standard deviation limit (3σ). If the limit is violated, a gross error is

Table 6. Comparison between predictions and measurements under different product loads.

Product load,(t/h)	$[C_{COOH}]$,(mol/T) (Measurement)	$[C_{COOH}]$,(mol/T) (Prediction)	Error percent of [COOH],(%)	DP (Measurement)	DP (Prediction)	Error percent of DP,(%)
6.496	639	652.1	2.10	16	16.1	0.67
6.608	684	667.8	2.31	16	16.2	1.44
6.717	690	709.9	2.84	15	15.2	1.11
6.717	655	698.4	6.56	20	19.6	2.17
6.853	623	660.2	6.04	19	18.8	0.94
6.896	652	690.7	5.86	16	16.1	0.63
6.897	712	735.7	3.35	18	17.8	1.00
7.046	650	639.6	1.68	17	17.1	0.60
7.138	646	682.9	5.77	15	15.0	0.27

When the industrial data are ready, MKEDA are used to estimate the parameters. The population size is set to 50 and each run continues until a maximum of 10,000 function evaluations is reached. The initial population is generated uniformly. Thirty percent of population is selected for model updating, and elitist strategy is used with the elitism individuals five. Figure 6 shows the Pareto solutions. In order to judge the quality of parameters, the simulation result coming from literature [6] is illustrated using the cross signal in Figure 6. It is obvious that the predictions using parameters in Yamada et al.[6] is far from the industrial data, and the parameters are not the non-dominated solution for this problem. For $[C_{COOH}]$ in the opposite direction with DP, there are more than one best solutions. Considering the measurement accuracy and the quality requirement, we choose a balance point of the Pareto set with the parameter values $[2.16 \times 10^6, 1.80 \times 10^6, 0.60 \times 10^6, 1.22 \times 10^6, 1.46 \times 10^6, 2.87 \times 10^6]$. Using these parameters to simulate the industrial PET reactor, the mean error of carboxyl end group content and DP is 13.60 and 0.89, respectively.

To test the performance of MKEDA, we tried other multi-objective evolutionary algorithms of NSGA-II, MPSO and MUMDA. The population size is set to 50, and each run continues until a maximum of 10,000 function evaluations is reached. The initial population is generated uniformly. Other specific parameters are the same as in Table 1. As all of these algorithms are heuristic methods, we have tested every algorithm for 30 times and the converged Pareto set are shown in Figure 7 comparing with MKEDA. It is shown that for this problem, the MKEDA can find better solutions with current parameter setting. It is because the probabilistic model can improve the solutions with the flexible kernel density estimation method comparing with Gaussian model. As to the NSGA-II and MPSO, the parameters setting are not appropriate for this optimization problem, so suitable parameters should be tried with trial and error method for many times to get better solutions. From this point of view, MKEDA performs more flexibility for the complex problems.

With the selected Pareto solution, the $[C_{COOH}]$ predictions are shown using the triangle signal in Figure 8 comparing with the measurement data. The horizontal coordinate is the industrial data, and the vertical coordinate is the prediction data. From the figure, it is obvious that the value is close to the diagonal. It means that the errors between the prediction and the real industrial measurements are small. Whereas, the simulation results from Yamada et al. [6] which is shown with the circle signal is far from the diagonal, which means that the error is out of the allowable accuracy and cannot be used for simulation of industrial process.

To validate the generalization ability of the model, more operating data are used for test. The production loads vary from 6.496 to 7.138 t/h, which cover most process conditions of the industrial reactors. The results are shown in Table 6. The predictions of the carboxyl end group concentration and the degree of polymerization agree with online measurements. Within the error percentage less than 7%, the model can be applied in the prediction of industrial process.

CONCLUSIONS

Different from the least squared method, which is widely used in the parameter estimation, this paper applied multi-objective optimization algorithm in estimating parameters estimation. From the point view of flexibility, a new method named MKEDA was proposed. Kernel density estimation is used instead of conventional Gaussian model, and several strategies including Pareto domination selection, and selection of kernel width and sampling method were adopted. MKEDA turned out to explore and exploit the search space more efficiently. Compared with other multi-objective algorithms including NSGA-II, MPSO and MUMDA, MKEDA could find better solutions in the kinetic parameters of PET reaction scheme. The new estimated parameters made the PET model reliable and robust in the prediction of product quality for industrial process.

REFERENCES

[1] T.-Y. Park, G.F. Froment. *Comput. Chem. Eng.*, **1998;** *22* (Supplement 1), S103–S110.

[2] V. Kumar, P. Balasubramanian. *Fuel*, **2009;** *88*(11), 2171–2180.

[3] J. Tao, N. Wang. *Comput. Chem. Eng.*, **2007;** *31*(12), 1602–1618.

[4] H. Zhang, G.P. Rangaiah. *Asia-Pac. J. Chem. Eng.*, **2011;** *6*(3), 379–390.

[5] A. Eftaxias, J. Font, A. Fortuny, A. Fabregat, F. Stüber. *Comput. Chem. Eng.*, **2002;** *26*(12), 1725–1733.

[6] T. Yamada, Y. Imamura, O. Makimura, H. Kamatani. *Polym. Eng. Sci.*, **1986;** *26*(10), 708–716.

[7] K. Ravindranath, R.A. Mashelkar. *Polym. Eng. Sci.*, **1982;** *22*(10), 610–618.

[8] S. Gupta. *J. Chem. Sci.*, **1983;** *92*(6), 673–680.

[9] V. Bhaskar, S.K. Gupta, A.K. Ray. *Polym. React. Eng.*, **2001;** *9*(2), 71–99.

[10] M. Ramteke, S.K. Gupta. *Int. J. Chem. React. Eng.*, **2011;** *9*, 55, 1–12.

[11] C.-K. Kang, B.C. Lee, D.W. Ihm. *J. Appl. Polym. Sci.*, **1996;** *60*(11), 2007–2015.

[12] N.N. Dorogov. *Theor. Found. Chem. Eng.*, **2000;** *34*(5), 485–490.

[13] H. Patel, G. Feix, R. Schomäcker. *Macromol. React. Eng.*, **2007;** *1*(4), 502–512.

[14] V. Bhaskar, S.K. Gupta, A.K. Ray. *AICHE J.*, **2000;** *46*(5), 1046–1058.

[15] V. Bhaskar, S.K. Gupta, A.K. Ray. *Comput. Chem. Eng.*, **2001;** *25*(2–3), 391–407.

[16] B.V. Babu, J.H.S. Mubeen, P.G. Chakole. *Mater. Manuf. Process.*, **2007;** *22*(5–6), 541–552.

[17] F. Manenti, M. Rovaglio. *Ind. Eng. Chem. Res.*, **2008;** *47*(1), 92–104.

[18] V. Bhaskar, S.K. Gupta, A.K. Ray. *Rev. Chem. Eng.*, **2000;** *16*(1), 1–54.

[19] S. Baluja. Population-based incremental learning: A method for integrating genetic search based function optimization and competitive learning; Technical Report CMU-CS-94-163; Carnegie Mellon University: Pittsburgh, PA, **1994.**

[20] G.R. Harik, F.G. Lobo, D.E. Goldberg. *IEEE Trans. Evol. Comput.*, **1999;** *3*(4), 287–297.

[21] H. Mühlenbein, G. Paaß. *Lect. Notes Comput. Sci.*, **1996;** *1141*, 178–187.

[22] S. Jeremy, B. De, L.I. Charles, P. Viola, MIMIC: finding optima by estimating probability densities. In *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, **1997.**

[23] S. Baluja, S. Davies. Combining multiple optimization with optimal dependency trees; Technical Report CMU-CS-97-157; Carnegie Mellon University, Pittsburgh, **1997.**

[24] M. Pelikan, H. Mühlenbein, The bivariate marginal distribution algorithm. In *Advances in soft computing: engineering design and manufacturing*, Springer-Verlag, London, **1999;** pp. 521–535.

[25] G. Harik, F. Lobo, K. Sastry, Linkage learning via probabilistic modeling in the extended compact genetic algorithm (ECGA). In *Studies in Computational Intelligence*, *Vol. 33*, Springer-Verlag, London, **2006;** pp. 39–61.

[26] M. Pelikan, D.E. Goldberg, E. Cantú-Paz, BOA: The Bayesian optimization algorithm. In *Genetic and Evolutionary Computation Conf.(GECCO-99)*, **1999.**

[27] H. Mühlenbein, T. Mahnig. *J. Comput. Inform. Technol.*, **1999;** *7*(1), 19–32.

[28] P. Larrañaga, R. Etxeberria, J.A. Lozano, J.M. Peña. Optimization in continuous domain by learning and simulation of Gaussian networks. In The Proceeding of the 2000 Genetic and Evolutionary Computation Conference Workshop Program, Las Vegas, Nevada, **2000.**

[29] M. Sebag, A. Ducoulombier. *Lect. Notes Comput. Sci.*, **1998;** *1498*, 418–427.

[30] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan. *IEEE Trans. Evol. Comput.*, **2002;** *6*(2), 182–197.

[31] X. Li. A non-dominated sorting particle swarm optimizer for multiobjective optimization. In *Genetic and Evolutionary Computation — GECCO 2003*, Vol *2723*, Springer, Berlin/Heidelberg, **2003;** 198–198.

[32] P.A.N. Bosman, D. Thierens. *Int. J. Approx. Reason.*, **2002;** *31*(3), 259–289.

[33] M. Laumanns, J. Ocenasek, Bayesian optimization algorithms for multi-objective optimization. In *Proceedings of the 7th International Conference on Parallel Problem Solving from Nature*, Springer-Verlag, London, **2002.**

[34] M. Costa, E. Minisci. *EMO 2003*, Faro Portugal, **2003;** pp. 282–294.

[35] S. Bureerat, K. Sriworamas, A. Saad, K. Dahal, M. Sarfraz, R. Roy, Population-based incremental learning for multiobjective optimisation. In *Advances in Soft Computing*, *Vol. 39*, Springer, Berlin / Heidelberg, London, **2007;** pp. 223–232.

[36] M. Pelikan, K. Sastry, D. Goldberg, Multiobjective estimation of distribution algorithms. In *Scalable Optimization via Probabilistic Modeling*, *Vol. 33*, Springer, Berlin / Heidelberg, **2006;** pp. 223–248.

[37] B. Li, R.T. Zhong, X.J. Wang, Z.Q. Zhuang In Proceeding of the 18th international conference on pattern recognition (ICPR 2006), Hong Kong, China, **2006;** 1192–1195.

[38] S. Tsutsui, M. Pelikan, D.E. Goldberg In Evolutionary algorithm using marginal histogram models in continuous domain, Proceeding of the 2001 Genetic and Evolutionary Computation Conference Workshop Program, **2001.**

[39] N. Ding, S. Zhou, Z. Sun. *J. Comput. Sci. Technol.*, **2008;** *23*(1), 35–43.

[40] P.A.N. Bosman, D. Thierens, Expanding from discrete to continuous estimation of distribution algorithms: The IDEA. In *Proceedings of the 6th International Conference on Parallel Problem Solving from Nature*, Springer-Verlag, London, **2000.**

[41] W. Hardle. *Smoothing Techniques with Implementation in S*, Springer-Verlag, New York, **1991.**

[42] J. Horn, N. Nafpliotis, D.E. Goldberg In Evolutionary computation, 1994. IEEE World Congress on Computational Intelligence, Proceedings of the First IEEE Conference on, **1994;** 82–87 *vol. 1*.

[43] S.A. Bhat, D.N. Saraf. *Ind. Eng. Chem. Res.*, **2004;** *43*(15), 4323–4336.