Multi-objective integrated solvent selection and solvent recycling under uncertainty using a new genetic algorithm

Weiyu Xu and Urmila M. Diwekar*

Center for Uncertain Systems,
Tools for Optimisation and Management (CUSTOM),
Vishwamitra Research Institute, Westmont, IL 60559, USA
E-mail: wxu6@uic.edu E-mail: urmila@vri-custom.org
*Corresponding author

Abstract: The optimal design of waste treatment processes always involves several objectives to be considered like cost and Environmental Impact (EI). Hence a multi-objective optimisation framework is required whose solution is not a single value but a Pareto set, which includes the alternatives representing potential compromise solutions among the objectives. Further, uncertainties are inherent in EI assessment, these uncertainties need to be propagated and analysed. In this paper, a multi-objective optimisation algorithm called MOEGA is developed. This new and efficient algorithm identifies more trade-off solutions (with and without uncertainties in EI) than before.

Keywords: multi-objective efficient genetic algorithm; MOEGA; Hammersley sequence sampling; HSS; uncertainty.

Reference to this paper should be made as follows: Xu, W. and Diwekar, U.M. (2007) 'Multi-objective integrated solvent selection and solvent recycling under uncertainty using a new genetic algorithm', *Int. J. Environment and Pollution*, Vol. 29, Nos. 1/2/3, pp.70–89.

Biographical notes: Weiyu Xu earned his Bachelors Degree in Chemical Engineering from Tianjin University, China and Masters Degree from the University of Illinois at Chicago. After completing his Masters, he pursued his PhD under Dr. Urmila Diwekar and his research dealt with environmentally benign product and process design under uncertainty. After successful completion of his PhD, he joined the 'Vishwamitra Research Institute' as a Post-Doctor Researcher.

Urmila M. Diwekar is the President and Founder of Vishwamitra Research Institute, a non-profit research organisation. From 2002–2004, she was a Full Professor in the Departments of Biochemical, and Industrial Engineering and the Institute for Environmental Science and Policy, University of Illinois. From 1991–2002 she was on the Faculty of the Carnegie Mellon University. She has worked extensively in the areas of simulation, design, optimisation, control, stochastic modelling, and synthesis of chemical processes. Uncertainties are inherent in real world processes. Recognising this fact, she started working in 1991 on efficient methods for uncertainty analysis, and optimisation under uncertainty. These led productive contributions in diverse fields and in 1999, she founded CUSTOM.

1 Introduction

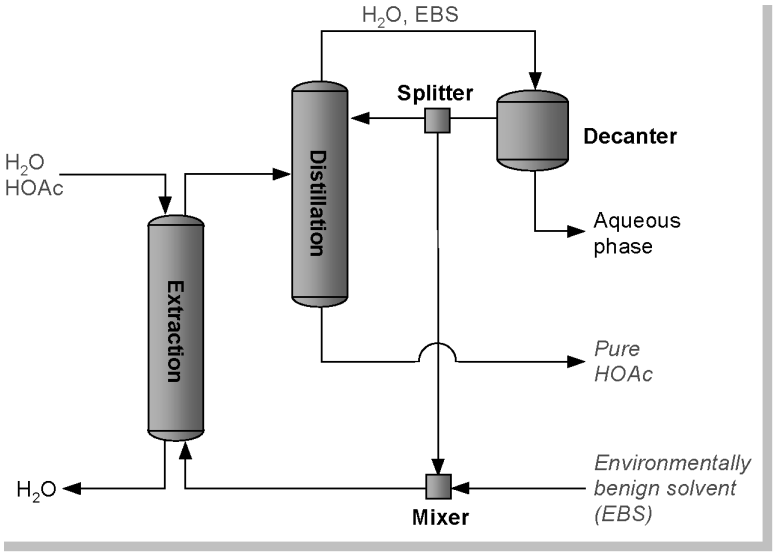
The treatment of acetic acid aqueous solution is an important industrial issue since there are various manufacturing processes which produce wastes or by-products containing acetic acid which is a valuable solvent as well as a pollutant to the environment when discharged without any treatment. Several competing processes have been proposed for recovering acetic acid from these aqueous solutions like extractive distillation (Lei et al., 2004) and heterogeneous azeotropic distillation (Downs and Siirola, 1997). Among these processes, heterogeneous azeotropic distillation is most suitable for the dilute acetic acid aqueous solution process. Figure 1 shows such a process configuration employed in Eastman Chemical. Co. (Downs and Siirola, 1997). In this process, a Environmentally Benign Solvent (EBS) is added to the mixture of acetic acid and water, then the mixture of water, acetic acid and EBS is fed to the distillation column. The overhead of the distillation column is the heterogeneous azeotropic mixture of solvent and acetic acid, which can be separated by a decanter. The bottom product of distillation column is highly pure acetic acid that can be reused. Though this process seems straightforward, it is very unstable due to the existence of an azeotrope. Any variation in the feed composition or the amount of extractant used would result in undesirable separation products. To increase the number of degrees of freedom for the azeotropic distillation column which is the most unstable component in this process, a new separation configuration is developed (Xu and Diwekar, 2005a) based on P-graph (Feng et al., 2003) technique, Residual Curve Maps (RCMs) analysis and optimisation technique that is shown in Figure 2. In this figure, the amount of recycle streams and feed flows fed to the distillation column to guarantee the purity of the bottom product, which ideally is pure acetic acid, is controlled by split fraction of splitters for feed flows and recycle streams. Unlike the earlier separation structure, in the new separation structure (Figure 2), the initial feed stream is split and one stream goes to the distillation column while the other part is sent to the extractor. The split fraction between these two operating units is determined by optimisation strategy. The distillate passes through a heat exchanger and is split again. One part goes to decanter while the other is recycled to the distillation column. This split ratio is again determined by the optimisation strategy. Unlike the conventional structure where only organic stream is recycled to the distillation column, in the new structure, both organic and aqueous streams can be recycled to the distillation column. The higher number of decision variables brought about by the new structure thus increases the overall process flexibility especially the flexibility of the azeotropic distillation column.

The EBS, which determines

- the shape of two-phase region
- the azeotrope composition
- the aqueous phase composition discharged from decanter

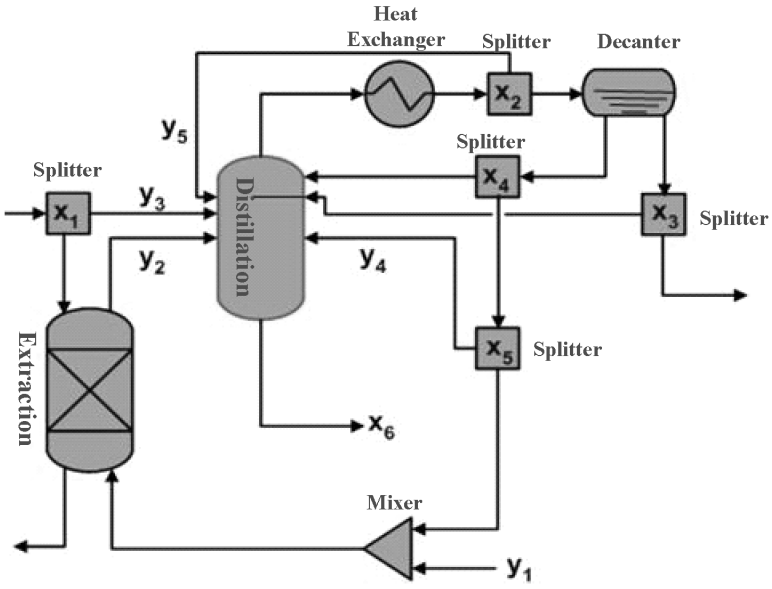
also has an effect on the separation performance.

Figure 1 The acetic acid recovery process using heterogeneous azeotropic distillation



Source: From Eastman Chemical Co.

Figure 2 The new separation configuration for acetic acid recovery



Thus attention should also be focused on the selection of solvents, which is accomplished by Computer-Aided Molecular Design (CAMD) (Kim and Diwekar, 2002), from which the molecules that satisfy property constraints and chemical feasibility constraints are identified. The solvents finalised in the CAMD step that also meet the requirements of the above separation process like the existence of azeotrope and two-liquid phase region are:

- ethyl acetate
- propyl acetate
- isopropyl acetate
- methyl propyl ketone
- methyl isopropyl ketone
- diethyl ketone
- methyl propionate.

These seven solvents are the potential EBS for the heterogeneous azeotropic distillation process in case of the acetic acid-water system.

In the integrated solvent selection and solvent recycling design, more than one objective needs to be taken into account. Several variants of Multi-Objective Genetic Algorithm (MOGA) have been widely used for the purpose of finding the Pareto set for the multi-objective optimisation problems. These variations include Vector Evaluated Genetic Algorithm (VEGA) (Schaffer, 1985), Niched Pareto Genetic Algorithm (NPGA) (Horn et al., 1994), Non-dominated Sorting Genetic Algorithm-II (NSGA-II) (Deb et al., 2002) and weighting method based MOGA (Leung and Wang, 2000; Ishibuchi and Murata, 1998). In VEGA, the population is divided into as many subpopulations as there are objectives. The individuals in each subpopulation are selected in terms of the corresponding objective for next generation. NPGA is extended from the traditional GA by adding two genetic operators:

- Pareto domination tournament
- sharing.

The basic idea is to initially pick two candidates and a comparison set at random. Then each of the candidates is compared to that comparison set. If one candidate is dominated by the comparison set while the other is not, the latter one is selected. If there is a tie between two candidates, irrespective of whether both are dominated or not by the comparison set, the less clustered candidate is selected by a sharing strategy to increase the population diversity. NSGA-II is implemented by non-dominated sorting. In NSGA-II, individuals are assigned to different Pareto fronts based on their priority ranking in each Pareto set. The overall Pareto set is ranked as the first Pareto front. The second Pareto front is obtained by removing the individuals in the first Pareto front and sorting the remaining individuals for non-dominated set, which is second Pareto front. This procedure is repeated until all the individuals have been assigned to Pareto fronts. A non-parameter sharing is also used in the selection of non-dominated solutions in the same Pareto front by computing shortest distance between adjacent individuals and the individual in a less crowded cluster is selected for the purpose of population diversity. The weighting method based Genetic Algorithm (GA) associates each population with a different set of weighting coefficients for diversifying the search directions and to optimise the weighted sum of the objective. By changing the weight of each objective, a set of Pareto solutions is found.

In this paper, a new variant of MOGA is developed using weighting method (Diwekar, 2003), where weights are generated by HSS technique (Kalagnanam and Diwekar, 1997; Diwekar and Kalagnanam, 1997) in order to increase the efficiency of the search. We apply this new algorithm to the integrated solvent selection and solvent recycling of acetic acid and water mixture. The objectives of the integrated solvent selection and process design are:

- maximise acetic acid recovery rate
- minimise the environment impact based on LC_{50}
- minimise the environment impact based on LD_{50}
- minimise the environment impact based on Bio-Concentration Factor (BCF)
- minimise the energy consumption
- maximise the process flexibility when facing the input flow fluctuation.

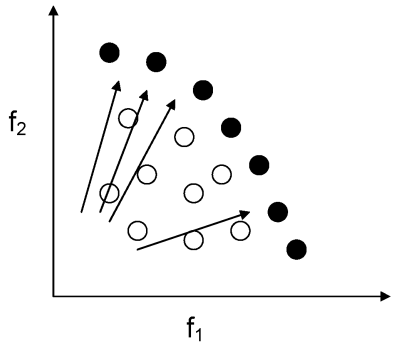
Another important issue is that the prediction models of LC_{50} and BCF are both regressed from the limited experimental data by minimising the difference between calculated value and experimental value and thus the errors are unavoidable. The optimisation without considering these errors would lead to sub-optimal solutions instead of the real optimum. Thus in this paper, the uncertainties of LC_{50} and BCF are characterised, quantified and propagated through the optimisation model to counter the effects of prediction errors.

This paper has three sections. In Section 2, a MOEGA is developed. In Section 3, a multi-objective integrated solvent selection and process design problem is formulated based on deterministic and stochastic frameworks and then solved using MOEGA developed in Section 2. The last part puts forth conclusions based on this work.

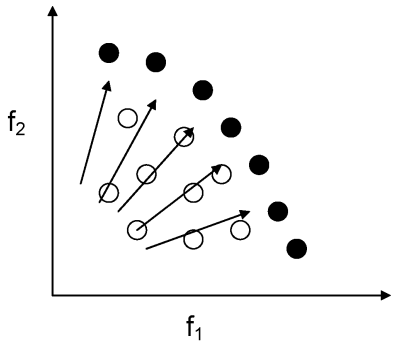
2 Multi-Objective Efficient Genetic Algorithm (MOEGA)

In the weighting method based MOGA, the weights play an important role in computing the Pareto set since every combination of weights for each objective represents a search direction. To identify more non-dominated solutions, more uniformly distributed search directions are preferred to the clustered search bias towards one region. A technique has been developed (Ishibuchi and Murata, 1998) which involves randomly assigning the weights to each objective of each population. The advantage of this method is it simplifies the decisions of assigning the weights, but the drawback is that there is no control on how to diversify the search directions. The random number generator is mostly developed based on Monte Carlo sampling, which can not distribute random samples uniformly on the search space. The consequence is that the search directions bias towards some directions as showed in Figure 3(a). Another strategy uses uniform design to select weights (Leung and Wang, 2000). The uniform design scatters the search directions uniformly towards the Pareto frontier, but it also complicates the weight selection and would consume considerable computation time.

Figure 3 The search directions for the Pareto front of a two-objective optimisation problem



(a)



(b)

To keep weight selection simple while make search directions as uniform as possible, we use an efficient sampling method, HSS, to assign the weights. HSS uses an optimal design scheme for placing N_{samp} points on a k-dimensional hypercube. This scheme ensures that the sample set is more representative of the population, showing better uniformity in the multi-dimensional uncertain surface compared to Monte Carlo. The uniformity property would spread out search directions towards Pareto frontier evenly as showed in Figure 3(b). At the same time, the strategies used in Efficient Genetic Algorithm (EGA) (Xu and Diwekar, 2005b, 2005c) are also employed in the development of MOEGA. Sharing is used to keep the population diversity. The MOEGA algorithm is summarised as follows:

At generation t = 0,

- *Use HSS method to generate initial population P(t), keep the k-dimensional uniformity property intact.*
- *Use HSS method to generate weights for each objective in each population.*
- *Evaluate P(t) by weighted sum objective function.*
- *Use sharing strategy to penalise clustered population.*
- *Generate tentative Pareto set.*

While termination condition is not satisfied,

- *Use HSS method to generate random moves in selection, crossover and mutation step by keeping the k-dimensional uniformity property intact.*
- *Recombine P(t) to generate C(t).*
- *Evaluate C(t) by the weighted sum objective function.*
- *Update the tentative Pareto set.*
- *Use sharing strategy to penalise clustered population.*
- *Select P(t + 1) from tentative Pareto set and C(t).*
- *Set P(t) = P(t + 1).*

Until Pareto set is stable or maximal generation is reached.

where $P(t)$ denotes the population at generation t. $C(t)$ denotes the candidates that are evaluated to yield $P(t + 1)$.

3 Multi-objective problem formulation for the integrated solvent selection and process design

The multi-objective integrated solvent selection and process design problem has been solved before (Xu and Diwekar, 2005a) with four objectives using combination of P-graph, RCM, and MOSA-NLP (SA-Constraint method) framework. In this section, we first solve the same problem configuration using MOEGA. Then two more objectives

- minimise the EI based on BCF
- minimise the reboiler energy consumption of distillation, are incorporated and a Pareto set for six objectives is generated.

Finally, the uncertainties in the prediction models of LC_{50} and BCF are characterised, quantified and propagated through the multi-objective optimisation framework. The stochastic Pareto set incorporating these uncertainties are generated using MOEGA and compared to the deterministic Pareto set.

3.1 Deterministic problem formulation with four objectives

The four objectives in this optimisation framework are

- maximise acetic acid recovery (Z_1)
- minimise EI based on LC_{50} (Z_2)
- minimise EI based on LD_{50} (Z_3)
- maximise the process flexibility (Z_4)

which is defined as the number of feasible solutions upon process uncertainty, which is expressed by feed variability. Unlike coupled simulated annealing and constraint method (SA-Constraint) that keeps the primary objective and transforms other objectives to the constraints, MOEGA assigns weights to each objective and optimises

them simultaneously. At the same time, to keep the objective functions on the same scale, normalisation is applied to all the objectives as showed in equation (1).

$$\overline{Z}_i = \frac{Z_i - Z_{i\min}}{Z_{i\max} - Z_{i\min}}. \tag{1}$$

Here, i is the index of objective, Z_i is the value of objective i, $Z_{i\min}$, $Z_{i\max}$ and $\overline{Z}_i$ are the minimum, maximum and normalised value of objective i at current generation. Then the optimisation problem is formulated as follows with the same constraints as previously solved using SA-Constraint method (Kim and Diwekar, 2002).

$$\text{Max} \sum_{i=1}^{n} w_i \times \overline{z}_i \tag{2}$$

$$\begin{aligned}
\text{s.t.}\quad & 1 \le y_1 \le 7, \quad 3 \le y_2 \le 10, \\
& 1 \le y_3 \le 10, \quad 1 \le y_4 \le 10, 1 \le y_5 \le 10, \\
& 0.0 \le x_1 \le 1.0, \quad 0.0 \le x_2 \le 1.0 \\
& 0.0 \le x_3 \le 1.0, \quad 0.0 \le x_4 \le 1.0 \\
& 0.0 \le x_5 \le 1.0, \quad 0.0 \le x_6 \le 1.0
\end{aligned}$$

$$\overline{Z}_1 = \text{normalised}\left\langle \frac{\text{Acetic acid in product}}{\text{Acetic acid in feed}} \right\rangle$$

$$\overline{Z}_2 = \text{normalised}\left\langle \frac{\sum F_{\text{solvent,out}}}{LC_{50,\text{solvent}}} + \frac{\sum F_{\text{HOAC,waste}}}{LC_{50,\text{HOAC}}} \right\rangle$$

$$\overline{Z}_3 = \text{normalised}\left\langle \frac{\sum F_{\text{solvent,out}}}{LD_{50,\text{solvent}}} + \frac{\sum F_{\text{HOAC,waste}}}{LD_{50,\text{HOAC}}} \right\rangle$$

$$\overline{Z}_4 = \text{normalised}\left\langle \frac{\text{feasible runs}}{\text{total runs}} \right\rangle, \quad y_i = \text{integer}$$

where x and y, which are also showed in Figure 2 are continuous and discrete decision variables respectively. The continuous decision variables are split fractions and bottom flow rate of distillation column. y_1 is solvent type, y_2 and y_3 are locations of feed streams, y_4 and y_5 are locations of recycle streams. n is the number of objectives, which is '4' in this formulation. The weight coefficients of Z_1 and Z_4 are positive, while weight coefficients of Z_2 and Z_3 are negative. This process is simulated using Aspen plus 12.1 simulation software and the MOEGA is embedded as a user block to perform the search for the Pareto set. The parameters of MOEGA are chosen from our previous paper (Xu and Diwekar, 2005b, 2005c). The population size used in MOEGA is 200; probabilities of crossover and mutation are 0.9 and 0.1; sharing parameter is 0.1.

The Pareto set computed by MOEGA is listed in Table 1, and the Pareto set found by the SA-Constraint method (Kim and Diwekar, 2002) is listed in Table 2. From the comparison, we can see that MOEGA has identified a higher number of potential solutions (Pareto solutions) than the SA-Constraint method. Ethyl acetate, which is one of the most commonly used extractants, appears in eight configurations in the Pareto set. These eight configurations all have favourable acetic acid recovery rates and also contain the highest value (0.9996) though it does not hold the largest distribution coefficient (a factor that affects the extraction defined as ratio of concentration of solute in extractive phase to the concentration of solute in raffinate phase) among the three selected solvents: ethyl acetate, isopropyl acetate and methyl propionate in the Pareto set. This shows that recovery rate is not only affected by extraction, but also is affected by the recovery process. Another observation is that though ethyl acetate (solvent index 1) has the largest LD_{50}, the lowest EI in terms of LD_{50} is held by isopropyl acetate (solvent index 3). The main reason is that the flow rates of solvent and acetic acid have a considerable effect on the EI value. The appearance of isopropyl acetate (solvent index 3) in the Pareto set is mainly due to its relatively high flexibility and lower EI in terms of LD_{50}. The disappearance of methyl isopropyl ketone from the Pareto set can be explained from the much larger EI value in terms of LD_{50}, which lowers the overall quality of the weighted sum objective value and consequently is not generated in the genetic operation.

Besides computing a more representative Pareto set, the MOEGA also substantially reduces the number of model calls (chemical process simulation). It has been found that for the four objective problem, the SA-constraint method required 32985 model calls to obtain the complete Pareto surface, as compared to the 15608 calls of MOEGA. From the comparison, it is evident that MOEGA saves 52.68% model calls.

Table 1 The Pareto set (four objectives) found by MOEGA with 5% feed flow variation (normal distribution)

Index	y_1	y_2	y_3	y_4	y_5	Z_1	Z_2	Z_3	Z_4
1	1	9	8	1	1	0.998312	0.7807	0.0311	0.23333
2	1	3	9	1	1	0.993087	0.8281	0.031	0.63333
3	1	9	9	4	9	0.995606	0.7231	0.0279	0.6
4	1	3	8	1	1	0.995844	0.7598	0.0292	0.5
5	1	5	2	3	1	0.999677	1.4034	0.0469	0.4
6	1	3	9	3	9	0.995716	0.9368	0.0333	0.56667
7	1	4	5	3	3	0.996498	0.6961	0.0267	0.16667
8	1	7	6	3	1	0.996165	0.7346	0.0284	0.36667
9	3	7	9	1	1	0.98537	1.107	0.0198	1
10	3	10	5	3	3	0.989137	0.9175	0.0161	0.8333
11	3	9	9	1	1	0.990936	0.9576	0.0163	0.7
12	7	7	4	5	2	0.989176	0.4089	0.0263	0.63333
13	7	9	6	5	3	0.998102	2.7331	0.3084	0.43333
14	7	5	8	3	2	0.99231	0.3411	0.02664	0.86667

Table 2 The Pareto set (four objectives) found by the SA-Constraint method with 5% feed flow variation (normal distribution)

Index	y_1	y_2	y_3	y_4	y_5	Z_1	Z_2	Z_3	Z_4
1	1	7	6	8	5	0.994	0.8315	0.0317	0.5667
2	1	10	4	10	3	0.9937	0.7714	0.0293	0.6
3	3	10	5	3	3	0.9891	0.9175	0.0161	0.8333
4	3	8	2	5	1	0.9727	1.2189	0.0252	1
5	5	9	8	2	5	0.9982	0.5308	1.1262	0.7
6	7	7	9	2	4	0.9842	0.5536	0.036	0.7
7	7	8	7	2	3	0.9908	0.4438	0.0328	0.5333

3.2 Deterministic problem formulation with six objectives

In this formulation, two more objectives are incorporated:

- minimise the EI based on BCF (Z_5)
- minimise the reboiler energy consumption of distillation (Z_6).

Besides LC_{50} and LD_{50}, which represent aquatic eco-toxicity and rodent toxicity (and possibly human toxicity) respectively, BCF is another important factor in risk assessment, which measures the long-term effects in individuals, populations and ecosystems. It is defined as the ratio of concentration of the chemical in an organism (such as fish) at steady state to the concentration of the chemical in the surrounding medium (such as water) during exposure period. In our design, the BCF values of seven solvent candidates are calculated from the model developed by Mackay (Devillers et al., 1996), which has the following formulation:

$$\log \mathrm{BCF} = 1.00 \log P - 1.32. \tag{3}$$

Here log P is the n-octanol/water partition coefficient. The log BCF values of seven solvent candidates and acetic acid are listed in Table 3.

Table 3 The log BCF value of seven solvent candidates and acetic acid

	Acetic acid	*Ethyl acetate*	*Propyl acetate*	*Isopropyl acetate*	*Methylpropyl ketone*	*Methyl isopropyl ketone*	*Diethyl ketone*	*Methyl propionate*
log BCF	–1.483	–0.59	–0.08	–0.04	–0.41	–0.48	–0.38	–0.58

Because the flows of pollutants should be minimised and the solvents should be safe, the definition of EI in terms of BCF is given as in equation (4).

$$\mathrm{EI} = \sum_{i}^{\text{stream}} \sum_{j}^{\text{pollutant}} \mathrm{Flow}_{ij} \times \mathrm{BCF}_{j}. \tag{4}$$

The primary motivation for employing heterogeneous azeotropic distillation is to avoid the considerable energy cost in case of simple distillation caused by the fact that all the water needs to be taken overhead due to it's higher volatility. In our new separation configuration as showed in Figure 2, a portion of feed flow is added to the distillation column to increase the stability of the process, but it goes against the motivation of employing heterogeneous azeotropic distillation which is to bring the least amount of water to the distillation column as possible. The extreme case is that all the feed flows are added to the distillation column and it turns to be a simple distillation. To strike a balance between the energy cost and process operability, an additional objective which is 'the reboiler energy consumption' is minimised while maximising the operability which is the flexibility in this case. The new optimisation model with additional objectives is formulated in the following equation (5).

$$\text{Max} \sum_{i=1}^{n} w_i \times \bar{z}_i \tag{5}$$

s.t. x_i and y_i $(i = 1,\ldots,6; j = 1,\ldots,5)$ use the same constraints as in equation (2)

$\bar{Z}_i (i = 1,\ldots,4)$ use the same formulas as in equation (2)

$$\bar{Z}_5 = \text{normalised}\left\langle \sum F_{\text{solvent,out}} \times \text{BCF}_{\text{solvent}} + \sum F_{\text{HOAc,waste}} \times \text{BCF}_{\text{HOAc}} \right\rangle$$

$$\bar{Z}_6 = \text{normalised}\left\langle \text{reboiler energy consumption} \right\rangle.$$

Here n is equal to 6 in this model, HOAc is acetic acid. The weight coefficients of Z_1 and Z_4 are positive, while weight coefficients of Z_2, Z_3, Z_5 and Z_6 are negative.

Totally, 46 solutions (Pareto set) are identified using MOEGA for this six-objective formulation (Table 4). Compared to the Pareto solutions found for 4-objective model, we can see that with the incorporation of more objectives, the number of trade-off solutions increases tremendously. In addition to ethyl acetate, isopropyl acetate and methyl propionate, methyl propyl ketone, methyl isopropyl ketone and diethyl ketone also appear for their smaller energy consumption and higher flexibility. Though methyl porpyl ketone and methyl isopropyl ketone have pretty large EI values in terms of LD_{50}, they still contribute to the Pareto set, which is not the case for the 4-objective model. This can be explained from the fact that with more objectives being incorporated, the effect of each objective to the weighted sum value becomes smaller. Thus a chance of methyl propyl ketone and methyl isopropyl ketone being in Pareto set is higher. The relationships between acetic acid recovery, flexibility and energy consumption are shown in Figure 4 as an example of trade-off between different objectives. From Figure 4, we can see that region A has advantages of high recovery rate and flexibility, while region B has the advantage of low energy consumption. The increase of recovery rate and flexibility is at the expense of operating cost.

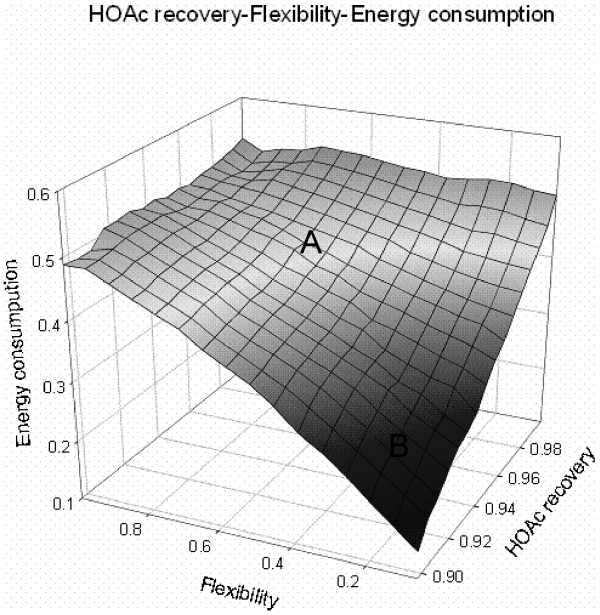
Table 4 The Pareto set (six objectives) found by MOEGA with 5% feed flow variation (normal distribution)

Index	y_1	y_2	y_3	y_4	y_5	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
1	1	7	1	1	1	0.990865	0.8888	0.0323	0.23333	0.37449	0.49161
2	1	5	2	8	1	0.987022	0.9718	0.0341	0.76667	0.36384	0.44751
3	1	9	8	1	1	0.998312	0.7807	0.0311	0.23333	0.43074	0.44923
4	1	3	9	1	1	0.993087	0.8281	0.031	0.63333	0.38125	0.51448
5	1	9	9	4	9	0.995606	0.7231	0.0279	0.6	0.36495	0.51925
6	1	3	8	1	1	0.995844	0.7598	0.0292	0.5	0.37887	0.51083
7	1	5	2	3	1	0.999677	1.4034	0.0469	0.4	0.44111	0.51141
8	1	5	2	2	1	0.992048	0.939	0.0347	0.53333	0.41817	0.49915
9	1	3	2	1	1	0.989766	0.9096	0.0328	0.7	0.37242	0.50031
10	1	4	2	1	1	0.990786	0.8926	0.0326	0.56667	0.38229	0.50097
11	1	3	9	3	9	0.995716	0.9368	0.0333	0.56667	0.36766	0.54177
12	1	5	9	2	1	0.992374	0.8953	0.0331	0.46667	0.39949	0.48578
13	1	3	1	9	9	0.975178	1.3099	0.043	0.8	0.38346	0.51886
14	1	4	3	6	2	0.990593	0.8886	0.032	0.46667	0.36471	0.51314
15	1	4	2	4	1	0.990359	0.8708	0.0312	0.36667	0.35096	0.50889
16	1	4	9	2	1	0.989533	0.8455	0.0308	0.63333	0.36101	0.47645
17	1	4	2	2	1	0.99039	0.8769	0.0319	0.76667	0.37045	0.51528
18	1	3	10	2	1	0.989838	0.9606	0.0345	0.56667	0.39124	0.48443
19	1	4	3	3	2	0.990812	0.8856	0.032	0.43333	0.36743	0.46258
20	1	4	5	3	3	0.996498	0.6961	0.0267	0.16667	0.34419	0.51106
21	1	7	6	3	1	0.996165	0.7346	0.0284	0.36667	0.37352	0.52964
22	1	5	2	2	4	0.993658	0.8902	0.0328	0.33333	0.39248	0.51021
23	3	7	9	1	1	0.98537	1.107	0.0198	1	0.64139	0.51987
24	3	4	3	9	1	0.897128	3.4468	0.0821	0.8	0.56085	0.48802
25	3	3	9	2	1	0.962231	1.8718	0.04	0.8	0.6255	0.47794
26	3	5	2	1	1	0.95665	1.7948	0.0386	0.76667	0.58503	0.46006
27	3	3	2	9	1	0.899349	3.3402	0.079	0.93333	0.5859	0.48897
28	3	9	9	1	1	0.990936	0.9576	0.0163	0.7	0.61128	0.55352
29	3	3	9	1	2	0.949595	1.7754	0.039	0.83333	0.51827	0.5313
30	3	5	8	3	4	0.974375	1.3231	0.0265	0.63333	0.56796	0.46627
31	3	3	4	4	1	0.936362	2.3008	0.0526	0.9	0.53075	0.50727
32	3	9	3	3	1	0.955663	1.5978	0.0344	0.76667	0.51687	0.51986
33	4	8	9	5	6	0.968355	1.27392	0.8435	0.83333	0.83132	0.48387
34	5	3	4	1	4	0.96906	1.025	1.0087	0.66667	0.50829	0.38574
35	5	4	3	1	1	0.980922	0.6885	0.9185	0.66667	0.45885	0.42196
36	5	3	1	1	1	0.959239	1.2705	0.9861	0.13333	0.5013	0.37368
37	5	3	1	3	1	0.953943	1.5944	1.0918	0.3	0.55815	0.37808
38	5	3	2	1	3	0.972518	0.9583	1.0038	0.33333	0.50484	0.3922

Table 4 The Pareto set (six objectives) found by MOEGA with 5% feed flow variation (normal distribution) (continued)

Index	y_1	y_2	y_3	y_4	y_5	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
39	5	3	4	1	2	0.985876	0.6152	1.0507	0.0.6667	0.52203	0.39085
40	6	3	3	4	1	0.899843	2.8661	0.1283	0.33333	0.58546	0.28988
41	6	7	2	2	1	0.934899	1.9845	0.1062	0.3	0.56537	0.37998
42	7	7	3	4	1	0.982205	0.6344	0.0339	0.53333	0.28365	0.40358
43	7	4	3	1	1	0.990307	0.433	0.0303	0.63333	0.30122	0.48698
44	7	3	1	2	1	0.975137	0.7931	0.0365	0.76667	0.26574	0.42591
45	7	7	4	5	2	0.989176	0.4089	0.0263	0.63333	0.25003	0.40679
46	7	6	9	7	5	0.99281	0.41703	0.0305	0.16667	0.31061	0.51852

Figure 4 Partial Pareto set of acetic acid recovery – process flexibility – energy consumption



3.3 Stochastic problem formulation with six objectives

3.3.1 Uncertainty analysis for BCF and LC_{50} model

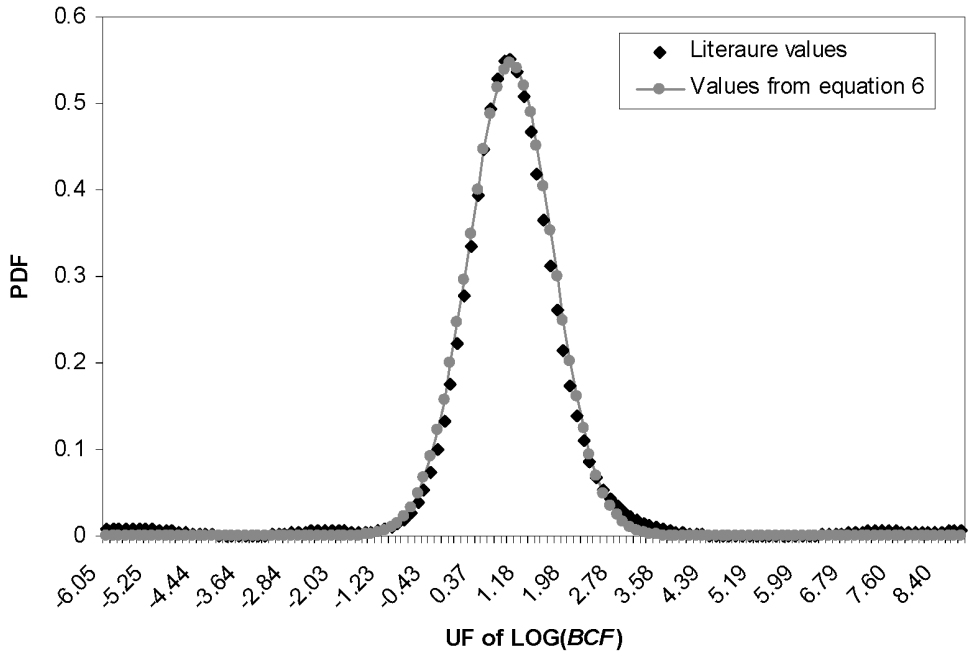
Both LC_{50} and BCF model are regressed from limited experimental data with the objective of minimising the differences between experimental and prediction values. Thus the prediction values are subjected to uncertainties (overestimation or underestimation). The omission of these uncertainties would lead to the fake 'optima' in the optimal design problem. So to analyse the effect of uncertainties on the design, an 'Uncertainty Factor' (UF) term is introduced, defined as the ratio of experimental value to the calculated value. A UF of 1 indicates an ideal case where the calculated value is exactly equal to the experimental value. In the BCF model, *n*-octanol/water partition coefficients (log *P*) which are subject to uncertainties also introduce estimation errors to

in the BCF prediction. These are included implicitly in the uncertainty analysis of BCF. The UF of BCF is defined in the following equation.

$$\mathrm{UF} = \frac{\log \mathrm{BCF_{exp}}}{\log \mathrm{BCF_{cal}}}. \tag{6}$$

The Probability Distribution Function (PDF) of a total of 178 UF data points (Devilles et al., 1996) is characterised and fitted to a normal distribution (Figure 5) with an arithmetic mean of 0.95 in UF and a standard deviation of 0.73 by the least squares method. From the distribution, we can see that PDF of UF is negatively shifted and thus the prediction results are tend to be overestimated. Because the log BCF values of all seven solvent candidates and acetic acid are negative, the overestimation of log BCF results in the underestimation of BCF. After incorporation of the UF, we would expect an increase of the EI in terms of BCF.

Figure 5 Probability distribution of UF for the log(BCF) from literature vs. theoretically calculated probability distribution using normal distribution function



The LC_{50} in this chapter is predicted by the group contribution approach (Gao et al., 1992) given by the following formula.

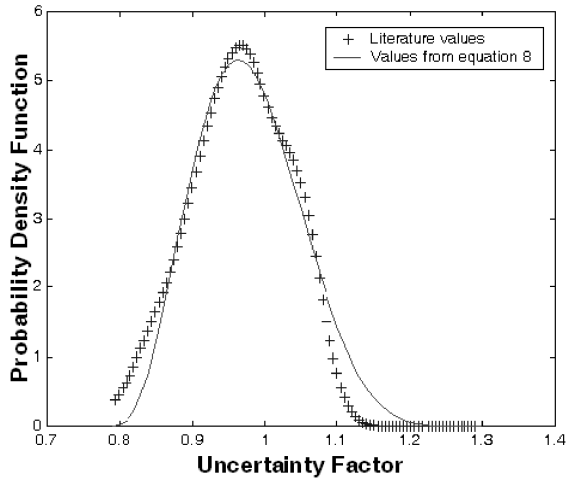
$$-\log \mathrm{LC_{50}} = \sum_{j=1}^{l} n_j \alpha_j \tag{7}$$

where n_j is the number of groups of type j, α_j is the contribution of groups of type j, l is the total number of groups in the compound. α_j are solved by regression analysis and a total of 16 groups are included. The UF of LC_{50} is defined in the following equation (8).

$$\mathrm{UF} = \frac{\mathrm{LC}_{50}^{\mathrm{exp}}}{\mathrm{LC}_{50}^{\mathrm{cal}}}. \tag{8}$$

The PDF of a total of 130 UF data points (Gao et al., 1992) are characterised and fitted in the beta distribution function (Figure 6). We used Beta distribution here because the distribution appears to be slightly skewed. Beta distribution is generally represented by four parameters like minimum, maximum (or mean and variance) and shape parameters P and Q. The two parameters P, Q for the distribution shown in Figure 6 are 4 and 6.86 respectively. The shift towards left of PDF leads to the underestimation of LC_{50} and consequently underestimates the EI based on LC_{50}. We would expect an increase of EI in terms of LC_{50} after propagating the uncertainties through the model.

Figure 6 Probability distribution of UFs for the LC_{50} from literature vs. theoretically calculated probability distribution using beta distribution function



3.3.2 Problem formulation

After employing the uncertainties, the integrated solvent selection and process design problem is formulated as follows.

$$\text{Max} \sum_{i=1}^{n} w_i \times \bar{z}_i \tag{9}$$

s.t. $\delta_1 \sim \text{beta}(4, 6.86)$ for the LC_{50}

$\delta_2 \sim N(0.95, 0.73)$ for the BCF

x_i and y_j $(i = 1,\ldots,6; j = 1,\ldots,5)$ use the same constraints as in equation (2)

$$\bar{Z}_1 = \text{normalised}\left\langle \frac{\text{Acetic acid in product}}{\text{Acetic acid in feed}} \right\rangle$$

$$\bar{Z}_2 = \text{normalised}\,\frac{1}{N}\sum_{j=1}^{N}\left\langle \frac{\sum F_{\text{solvent,out}}}{\delta_1^j LC_{50,\text{solvent}}} + \frac{\sum F_{\text{HOAC,waste}}}{\delta_1^j LC_{50,\text{HOAC}}} \right\rangle$$

$$\bar{Z}_3 = \text{normalised}\left\langle \frac{\sum F_{\text{solvent,out}}}{LD_{50,\text{solvent}}} + \frac{\sum F_{\text{HOAC,waste}}}{LD_{50,\text{HOAC}}} \right\rangle$$

$$\bar{Z}_5 = \text{normalised}\frac{1}{N}\sum_{j=1}^{N}\left\langle \sum F_{\text{solvent,out}} \times 10^{\delta_2^j \log(\text{BCF}_{\text{solvent}})} + \sum F_{\text{HOAc,waste}} \times 10^{\delta_2^j \log(\text{BCF}_{\text{HOAc}})} \right\rangle$$

$$\bar{Z}_6 = \text{normalised}\left\langle \text{reboiler energy consumption} \right\rangle$$

where δ is an uncertain parameter of UF and is imposed on the log(BCF) and LC_{50}. N is the number of uncertain samples. In this model, we take a constant number of samples (30) that balances the accuracy of representing the UF distributions and calculation efficiency. The other variables have the same meaning as in equation (2).

The MOEGA is applied to this model. Figure 7 represents the generalised solution procedure where the ASPEN model with all the configuration is augmented with the sampling loop accounting for feed fluctuations, an additional sampling loop represents the uncertainties in LC_{50} and BCF parameters, and the MOEGA optimises iteratively with these two sampling loops. The results of this analysis are presented in Table 5. The total number of Pareto solutions has been decreased to 30. As expected, the EIs in terms of LC_{50} (Z_2) and BCF (Z_5) both have increased after uncertainty propagation. From Table 4, we can see that most of the Z_2 and Z_5 values are below 1. But in Table 5, most Z_2 and Z_5 values are well above 1. Ethyl acetate (solvent index 1), the most promising solvent in the deterministic case which featured 22 times in Pareto solutions in a total of 46, only contributes to five configurations in the stochastic case. Methyl propyl ketone (solvent index 4), methyl isopropyl ketone (solvent index 5) and diethyl ketone (solvent index 6) that do not have an important portion of contribution to the deterministic Pareto set, have more presences and show they are also very promising solvent candidates.

Figure 7 Optimisation procedure of MOEGA

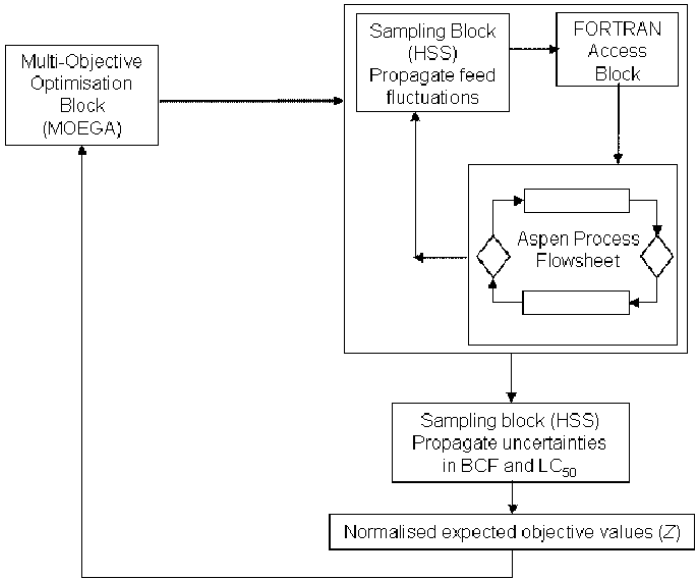


Table 5 The Pareto set (six objectives) under uncertainty found by MOEGA with 5% feed flow variation (normal distribution)

Index	y_1	y_2	y_3	y_4	y_5	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
1	1	9	1	1	1	0.979924	1.1779	1.2905	0.16667	1.99788	0.42415
2	1	9	9	3	9	0.992165	0.9197	0.031	0.6	1.15147	0.5073
3	1	3	9	2	2	0.90977	3.4981	0.0718	0.53333	2.52962	0.46949
4	1	7	6	3	1	0.996165	1.7955	0.0284	0.36667	2.41655	0.52964
5	1	4	9	9	3	0.988596	1.962	0.3824	0.7	2.49878	0.55221
6	3	3	2	1	5	0.988082	0.6779	1.2287	0.5	1.59339	0.45807
7	3	3	2	8	1	0.996267	0.3575	0.0314	0.6	1.06149	0.47731
8	3	6	4	6	4	0.988244	2.2439	0.1186	0.16667	2.62995	0.41542
9	3	7	4	3	2	0.986184	0.8659	0.1963	0.76667	1.16015	0.48039
10	3	10	7	8	1	0.992716	1.6702	0.1595	0.96667	2.24376	0.49264
11	3	5	8	1	3	0.988652	1.0308	0.0339	0.7	1.24783	0.52488
12	4	6	2	2	7	0.939673	2.9319	0.0589	1	2.1291	0.52528
13	4	9	2	7	4	0.944939	2.5525	0.9774	0.86667	1.87501	0.47752
14	4	8	9	5	6	0.968355	2.8703	0.8435	0.83333	3.02499	0.48387
15	4	7	7	5	2	0.955231	1.5591	0.9128	0.76667	2.16698	0.46419
16	5	3	4	1	4	0.96906	0.9167	1.0087	0.66667	1.16142	0.38574
17	5	3	2	1	4	0.996188	0.8389	0.03	0.16667	1.13214	0.51726
18	5	7	1	5	1	0.942488	2.2678	0.0449	0.86667	1.65183	0.48702
19	5	5	4	2	1	0.959231	2.2547	0.122	0.3	2.68603	0.44194
20	5	3	1	5	1	0.964096	3.1088	0.1387	0.4	3.25823	0.41252
21	5	7	2	4	6	0.970412	1.3365	1.5438	0.2	2.33942	0.44127
22	6	7	2	2	1	0.934899	0.1902	0.1062	0.3	0.89573	0.37998
23	6	4	2	2	2	0.992118	0.8696	0.0294	0.46667	1.09291	0.53507
24	6	3	3	4	1	0.899843	1.6713	0.1283	0.33333	2.34471	0.28988
25	6	4	4	1	1	0.962816	1.6547	0.1009	0.26667	2.14181	0.33214
26	6	8	4	4	3	0.99281	0.4611	0.0305	0.16667	1.02288	0.51852
27	6	6	2	5	8	0.969109	3.606	0.1609	0.26667	3.78027	0.39642
28	6	7	9	2	4	0.977091	1.836	1.0072	0.1	2.63524	0.35031
29	7	7	8	1	1	0.990567	0.9293	0.043	0.66667	1.42279	0.50127
30	7	9	6	5	3	0.998102	0.8028	0.3084	0.43333	1.09667	0.54543

4 Conclusions

In this paper, a MOEGA is developed by exploiting the *k*-dimensional uniformity property of HSS method. The process of recycling of acetic acid from its aqueous solution is formulated as a multi-objective problem with the consideration of EI and process operability. The recycle strategy developed from the P-graph technique is simulated in Aspen plus 12.1 and the optimisation is achieved by implementing MOEGA

as a user block which enables the simulation and optimisation to proceed simultaneously. A higher number of Pareto solutions are identified with this algorithm compared to the simulated annealing based SA-Constraint method. To obtain a more comprehensive evaluation of this recycle strategy, two more objectives are incorporated into the multi-objective model

- EI based on BCF
- energy consumption.

Another aspect of this paper is to characterise and quantify the uncertainties in the prediction model of LC_{50} and BCF to counteract the deviation caused by estimation errors. By propagating these uncertainties through the optimisation model and comparing the stochastic and Pareto surfaces, it was found that these uncertainties have a considerable effect on the optimal designs.

Acknowledgements

The financial supports of the EPA's STAR program through the grant #R38128 and NSF through the grant #CTS-0353088 are gratefully acknowledged.

References

Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T. (2002) 'A fast and elitist multiobjective genetic algorithm: NSGA-II', *IEEE Transactions on Evolutionary Computation*, Vol. 6, No. 2, pp.182–197.

Devillers, J., Bintein, S. and Domine, D. (1996) 'Comparison of BCF models based on Log P', *Chemosphere*, Vol. 33, pp.1047–1065.

Diwekar, U.M. (2003) 'Introduction to applied optimization', *Applied Optimization*, Vol. 80, Kulwer Academic Publishers, Massachusetts.

Diwekar, U.M. and Kalagnanam, J.R. (1997) 'Efficiency sampling technique for optimization under uncertainty', *AIChE J*., Vol. 43, pp.440–447.

Downs, J.J. and Siirila, J.J. (1997) 'Challenges of integrated process and control systems design for operability', *AIChE Annual Meeting*, Los Angeles, CA, p.189a.

Feng, G., Fan, L.T., Seib, P.A., Bertok, B., Kalotai, L. and Friedler, F. (2003) 'Graph-theoretic method for the algorithmic synthesis of azeotropic-distillation systems', *Ind. Eng. Chem. Res*., Vol. 42, pp.3602–3611.

Gao, C., Govind, R. and Tabak, H. (1992) *Application of the Group Contribution Method for Predicting the Toxicity of Organic Chemicals*, HH Environmental Toxicology and Chemistry ETOCDK, Vol. 11, No. 5, pp.631–636.

Horn, J., Nafpliotis, N. and Goldberg, D.E. (1994) 'A Niched Pareto Genetic Algorithm for multi-objective optimization', *Proc. 1st IEEE Int. Conf. Evolutionary Comptat*., pp.82–87.

Ishibuchi, H. and Murata, T. (1998) 'A multi-objective genetic local search algorithm and its application to flowshop scheduling', *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, Vol. 28, No. 3, pp.392–403.

Kalagnanam, J.R. and Diwekar, U.M. (1997) 'An efficient sampling technique for off-line quality control', *Technometrics*, Vol. 39, pp.308–319.

Kim, K.J. and Diwekar, U.M. (2002) 'Efficient combinatorial optimization under uncertainty. 2. Application to stochastic solvent selection', *Ind. Eng. Chem. Res*., Vol. 41, pp.1285–1296.

Lei, Z., Li, C., Li, Y. and Chen, B. (2004) 'Separation of acetic acid and water by complex extractive distillation', *Separation and Purification Technology*, Vol. 36, pp.131–138.

Leung, Y. and Wang, Y. (2000) 'Multiobjective programming using uniform design and Genetic Algorithm', IEEE *Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, Vol. 30, No. 3, pp.293–304.

Schaffer, J.D. (1985) 'Multi-objective optimization with Vector Evaluated Genetic Algorithms', *Proc. 1st Int. Conf. Genetic Algorithms*, pp.93–100.

Xu, W. and Diwekar, U.M. (2005a) 'Environmentally friendly heterogeneous azeotropic distillation system design: integration of EBS selection and IPS recycling', *Industrial and Engineering Chemistry Research*, Vol. 44, pp.4061–4067.

Xu, W. and Diwekar, U.M. (2005b) 'Improved Genetic Algorithms for deterministic and optimization under uncertainty-Part I: algorithm development', *Industrial and Engineering Chemistry Research*, Vol. 44, pp.7132–7137.

Xu, W. and Diwekar, U.M. (2005c) 'Improved Genetic Algorithms for deterministic and optimization under uncertainty-Part II: solvent selection under uncertainty', *Industrial and Engineering Chemistry Research*, Vol. 44, pp.7138–7146.

Nomenclature

BCF	Bio-Concentration Factor
CAMD	Computer-Aided Molecular Design
EBS	Environmentally Benign Solvent
EGA	Efficient Genetic Algorithm
EI	Environmental Impact
GA	Genetic Algorithm
HSS	Hammersley Sequence Sampling
LC_{50}	Lethal Concentration of 50% test population
LD_{50}	Lethal Dose of 50% test population
MOEGA	Multi-Objective Efficient Genetic Algorithm
MOGA	Multi-Objective Genetic Algorithm
MOSA-NLP	Multi-Objective Simulated Annealing – Non-Linear Programming
NPGA	Niched Pareto Genetic Algorithm
NSGA-II	Non-dominated Sorting Genetic Algorithm-II
P-graph	Process graph
RCMs	Residual Curve Maps
RHS	Right Hand Side
SA	Simulated Annealing
UF	Uncertainty Factor
VEGA	Vector Evaluated Genetic Algorithm
x	Splitter fractions
y	The feed or recycle streams locations in azeotropic distillation column
Z_1	Acetic acid recovery rate
Z_2	EI based on LC_{50}
Z_3	EI based on LD_{50}

Z_4	Process flexibility
Z_5	EI based on BCF
Z_6	Reboiler energy consumption of distillation
Z_{min}	Minimal objective value
Z_{max}	Maximal objective value
$\overline{Z}$	Normalised objective value