Integrated Solvent Selection and Recycling for Continuous Processes

Ki-Joo Kim and Urmila M. Diwekar*

CUSTOM (Center for Uncertain Systems: Tools for Optimization and Management) and Civil & Environmental Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213

This paper presents a real world case study where a multiobjective programming (MOP) framework under uncertainty is used for *simultaneous* integration of environmentally benign solvent (EBS) selection and in-process solvent (IPS) recycling. At the EBS selection level within this framework, the Hammersley stochastic annealing algorithm is applied to design candidate EBSs under uncertainty. This algorithm can efficiently optimize stochastic combinatorial optimization problems and generate a different set of candidate EBSs from that of the deterministic EBS selection model. At the IPS recycling level, Aspen Plus with a nonlinear programming technique is used to optimize the acetic acid recovery process. Then, these EBS selection and IPS recycling models are integrated under the MOP framework. At the MOP level, an efficient constraint MOP algorithm is employed to evenly approximate the Pareto solution surface (i.e., tradeoff surface). Four objectives—acetic acid recovery, process flexibility, and two environmental impacts based on LC_{50} and LD_{50}—are evaluated. The resulting MOP framework provides very distinctive chemical and process design alternatives (i.e., Pareto optimal solutions), and uncertainties in this framework significantly affect the size and shape of the Pareto set. This novel MOP framework can be applied to any large-scale stochastic mixed-integer nonlinear optimization problems because this framework is computationally efficient even in the case of combinatorial explosion and uncertainty inclusion.

1. Introduction

Separation processes not only are vital for isolating and purifying valuable products but also are crucial for removing toxic and hazardous substances from waste streams emitted to the environment. Among the various separation processes in chemical process industries, distillation process is most commonly used. Figure 1 shows one example of separation processes using extraction and distillation for acetic acid (HOAc) separation from water and its recycling.[1] HOAc, an *in-process* solvent (IPS), is a valuable chemical but also a pollutant when released to the environment. HOAc can be directly separated from water in a single distillation column; however, this requires a very large number of equilibrium stages because of close boiling temperatures of water and HOAc (100 °C vs 118 °C), resulting in high capital and operating costs. Instead of using a single distillation column, this separation, in practice, consists of an extraction column followed by a distillation column. An aqueous stream containing HOAc enters the extraction column in which an *environmentally benign* solvent (EBS) extracts HOAc from the aqueous mixture. The extract is then supplied to the azeotropic distillation column where the bottom product is pure HOAc and the top product is a heterogeneous water−EBS azeotrope. The pure HOAc product is recycled to upstream processes, whereas the azeotropic mixture is condensed and then decanted. The organic phase from the decanter is recycled to the two columns, while the aqueous phase goes to the wastewater treatment facility. A *split fraction*, which is an important operating variable, represents how much of the organic phase from the decanter recycles to the distillation column. Because EBS can be lost at the effluent streams, fresh makeup EBS must be added continuously. This HOAc recovery process is built in Aspen Plus, a steady-state simulator, where ethyl acetate (EtOAc) is currently used as an extracting solvent.

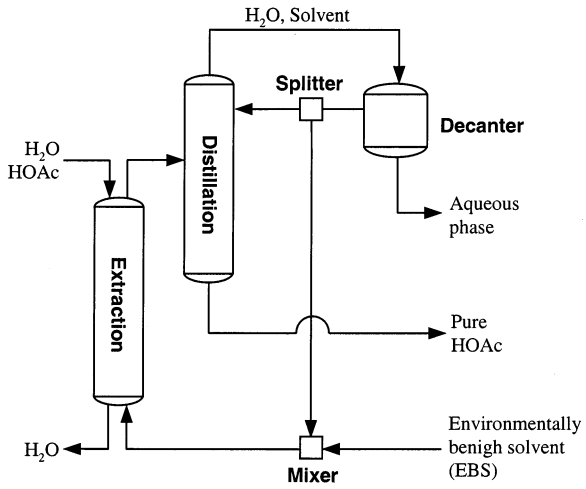

Figure 1. HOAc extraction process as a case study (from Eastman Chemical Co.).

This process seems to be simple and easy to operate. However, this process has several challenging problems. It can be easily shown that this configuration has no degrees of freedom for improving the process performance and flexibility. This is mainly due to the azeotropic constraint, the type of extracting solvent used, and the steady-state nature of the process. Therefore,

* To whom correspondence should be addressed. E-mail: urmila@cmu.edu. Phone: +1 412 268 3003. Fax: +1 412 268 7813.

10.1021/ie010777g CCC: $22.00
Published on Web 02/01/2002

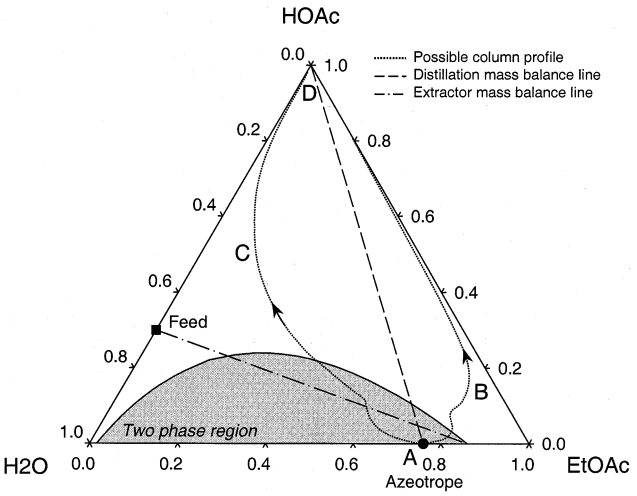

Figure 2. Ternary phase diagram of $H_2O-HOAc-EtOAc$ (the closed square is a feed point, and the closed circle is an azeotropic point).

variations in the feed condition can lead to severe process instability that will cause great economic loss and large environmental impacts (EI). A ternary phase diagram of water−HOAc−EtOAc is shown in Figure 2, in which the gray region is a two-phase region. A small variation in the feed composition, which is not welcomed but is common in real problems, can disturb the performance and stability of the current process. The composition in the extract stream, which becomes feed to the distillation column, is affected through the extraction mass balance line, and the bottom composition point (D) of the distillation column is severely affected by this feed variation. Feed composition variation can also change the locus of distillation profiles, resulting in completely unexpected products. If the distillation profile follows the track A−C−D, this results in a bottom product with low water purity, which may be acceptable. However, if the profile follows the track A−B−D, then the bottom product has a very low purity in the extracting agent composition and is not acceptable as a final product. Further, although the distillation profiles move toward the vertex of HOAc in the ternary diagram, the bottom composition point (D) may not be close to the vertex of HOAc, resulting in low HOAc purity (say, 80%), which is even worse.

Figure 3 shows a sensitivity analysis of the HOAc recovery yield with respect to the bottom flow rate and the split fraction when EtOAc is used as an extracting agent. The dark regions (mostly on the left side) represent low HOAc recovery yields, while the bright regions represent high HOAc recovery yields. The feasible region for high HOAc recovery is, unfortunately, very limited and sporadic and is very sensitive to operating and design conditions. This significantly affects the convergence and size of the optimal solutions. The feasible region is also affected by feed composition variations and extracting agent types. For example, process flexibility, defined as the number of feasible designs to the total number of designs, of three extracting agents—EtOAc, propyl acetate, and methyl propionate—at a particular operating condition is 50%, 70%, and 53%, respectively. Thus, including the EBS selection along with IPS recycling is essential for improving the stability and feasibility of this separation process.

The goals of this process design are to achieve a high HOAc recovery yield, a high process flexibility, and low

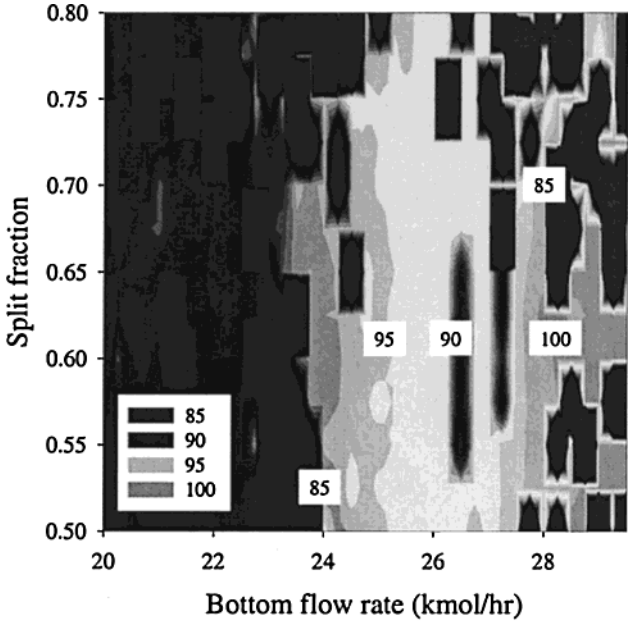

Figure 3. Sensitivity analysis of HOAc recovery rate.

EI. To evaluate the overall economic and environmental performances of this process, a simultaneous integration of EBS selection and IPS recycling should be considered. EBS selection is an approach used to generate candidate solvent molecules that have desirable physical, chemical, and environmental properties. Computer-aided molecular design (CAMD)[2,3] is commonly used for EBS selection. CAMD, based on the reverse use of group contribution methods, can automatically generate promising solvent molecules from their fundamental building blocks or groups.

In recent years, researchers have realized the importance of including chemical synthesis in process design. Chemical synthesis in this paper refers to EBS selection, while process design refers to IPS recycling process design. Buxton et al.[4] presented an integrated CAMD model in absorption processes. They formulated a mixed-integer nonlinear programming problem to minimize the global environmental impact (or cost). Hostrup et al.[5] also presented a similar integrated framework for coupled chemical and process synthesis. They presented a hybrid method of mathematical modeling with heuristic approaches in the superstructure formulation and solved the optimization problem to achieve minimum cost or energy consumption while satisfying environmental and process constraints.

However, this integration poses the significant problem of a combinatorial explosion of chemical and process design alternatives, and uncertainties over the chemical synthesis, process synthesis, and simulation stages add additional complexities to this problem. In addition, because the integrated framework needs to consider multiple objectives such as profitability, energy consumption, and EI, this coupled framework requires a multiobjective programming (MOP) technique. Recently, many researchers have focused on the applications of MOP mainly to design for the environmental problems in chemical engineering,[6−8] and Bhaskar et al.[9] published a good review paper on this topic, including various MOP solution techniques such as the constraint method, goal programming, and genetic algorithm.

This paper presents a novel MOP framework for EBS selection and IPS recycling under uncertainty. For the

EBS selection level in the framework, we used the Hammersley stochastic annealing (HSTA) algorithm,[10] an efficient discrete optimization algorithm under uncertainty. For the improvement of the MOP level, we applied an efficient constraint MOP algorithm based on the work of Fu and Diwekar.[11] For the sampling level, an efficient sampling technique, the Hammersley sequence sampling (HSS) technique, is applied to generate uncertain samples. By using these efficient methods at three levels, this novel MOP framework can evenly approximate the Pareto optimal surface and reduce the total optimization time. Thus, this proposed framework presents a useful tool for any multiobjective optimization problems due to features such as computational efficiency and the systematic inclusion of uncertainty.

This paper has three main sections: the MOP framework, EBS selection under uncertainty, and coupled EBS selection and IPS recycling as an MOP problem. The efficient MOP framework is explained in the next section for the simultaneous integration of chemical synthesis and process design. The section of EBS selection under uncertainty describes EBS selection criteria, a stochastic CAMD model, and distinctive results. The MOP problem section formulates an MOP problem under uncertainty and provides Pareto optimal solutions at two different uncertainty cases. The last section concludes this paper.

2. MOP Framework

The EBS selection and IPS recycling problem shown in Figure 1 is a MOP problem, which involves multiple objectives and tradeoffs between optimal solutions. The objectives of this simultaneous integration are to maximize HOAc recovery, maximize process flexibility, minimize environmental impact based on LC_{50} (lethal concentration at 50% mortality), and minimize environmental impact based on LD_{50} (lethal dose at 50% mortality). Process flexibility in this paper is defined as the number of feasible solutions in the uncertain region of feed variability. Because the flows of pollutants should be minimized and the solvents should be safe, the environmental impact defined in terms of LC_{50} and LD_{50} is given as

$$\mathrm{EI} = \sum_{i}^{\text{stream}} \sum_{j}^{\text{pollutant}} \frac{\mathrm{Flow}_{ij}}{\mathrm{LC}_{50,j}\,(\text{or } \mathrm{LD}_{50,j})} \qquad (1)$$

The environmental impact based on fathead minnow LC_{50} (mg/L) represents aquatic ecotoxicity, while the one based on oral rat LD_{50} (mg/kg) represents rodent toxicity (and possibly human toxicity).

MOP can be thought of as a set of methodologies for generating a preferred solution or range of efficient solutions to a decision problem[12] and can provide the least objective conflict. The preferred set is also known as the nondominated set or the *Pareto* set, which is a collection of alternatives that represent potential compromise solutions among the objectives. More specifically, it is a set of solutions that are superior to the rest of the solutions in the objective space but are inferior to other solutions in the space in one or more objectives.[13]

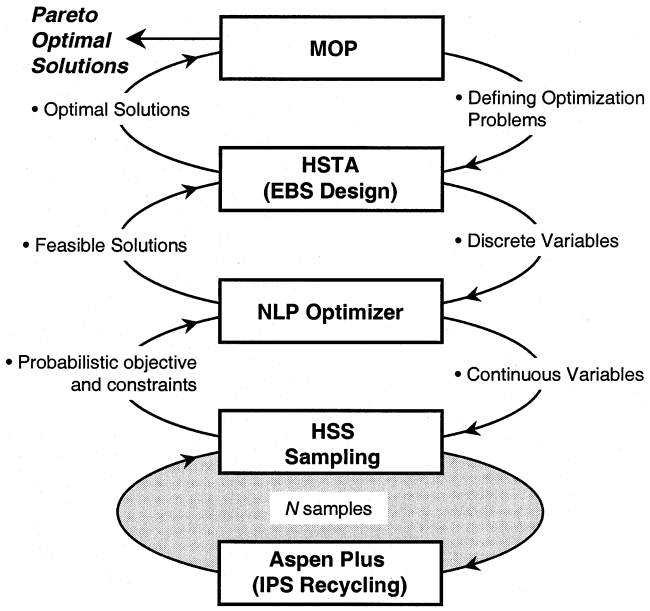

Figure 4. MOP framework for EBS selection and IPS recycling.

To solve multiobjective optimization problems, it is necessary to develop a complete multiobjective surface so that the full range of alternatives would be known and the tradeoffs among the objectives implied by each alternative would be understood. This involves finding the Pareto optimal solutions—*not a single solution*—from a very large number of design alternatives, such that no one solution dominates any of the others in the population.

There are a large array of analytical techniques to solve this MOP problem, and in this paper we apply an improved constraint MOP method, which is a pure algorithmic approach. This method can pick one of the objectives to minimize arbitrarily while the remaining others are turned into inequality constraints, with parametric right-hand sides, L_k. The problem takes on the following form:

$$\min Z_j \qquad (2)$$

$$\text{s.t.} \quad h(x,y) = 0$$

$$g(x,y) \le 0$$

$$Z_k \le L_k,\ k = 1, \ldots, j-1, j+1, \ldots, p$$

where Z_j is the chosen jth objective that we wish to optimize and p is the total number of objective functions. Solving repeatedly for different values of L_k leads to the Pareto set, and this approach is equivalent to calculating an integral over the space of objectives.

Because of the large size of the Pareto set and the iterative nature of this solution technique, a multiobjective optimization problem requires an efficient MOP algorithm to obtain the Pareto set within a reasonable time scale. Figure 4 represents the improved MOP framework, which is a multiobjective stochastic annealing nonlinear programming framework. Detailed descriptions and efficiency improvements of this framework are explained in the following paragraph.

Level 1: Aspen Plus[14] for IPS Recycling. Aspen Plus, a steady-state chemical process simulator, is used to build this case study.

Level 2: HSS. This sampling loop is needed for uncertainty analysis. The diverse nature of uncertainty, such as estimation errors and process variations, can be specified in terms of probability distributions and can be performed by sampling the distributions in an iterative manner. Of several possible sampling techniques, the HSS technique is applied because it is known that the HSS technique is at least 3−100 times faster than other current state-of-the-art techniques such as Latin hypercube sampling and Monte Carlo sampling.[15]

Level 3: Nonlinear Programming Optimizer for Continuous Optimization. The successive quadratic programming (SQP) method is used because it requires far fewer function and gradient evaluations than other methods for constrained optimization and it does not need feasible points at intermediate iterations. Both of these properties make SQP one of the most promising techniques for problems dealing with nonlinear constraint optimization, like process simulations.

Level 4: HSTA for EBS Design. This level provides an efficient interaction between discrete optimization methods and sampling techniques for uncertainty analysis. Simulated annealing, developed by Kirkpatrick et al.,[16] is a good candidate for discrete optimization problems because it can be applied to highly nonconvex systems. Because of a combinatorial explosion of the EBS selection problem, an efficient simulated annealing is developed by utilizing the uniformity property of the HSS technique in order to minimize the number of combinations of discrete decisions. Based on this efficient simulated annealing algorithm and the stochastic annealing algorithm,[17,18] the HSTA algorithm[10] is developed for efficient discrete optimization under uncertainty. In this algorithm, the HSS technique is also exploited to reduce the number of uncertain samples by tightening their error bandwidths. It is observed that the computational efficiency of the HSTA algorithm is 64% greater than that of the stochastic annealing algorithm and 99% greater than that of a conventional discrete optimization under uncertainty with a fixed number of samples for uncertain variables.[10] Hence, this algorithm holds a great promise for complex large-scale problems involving discrete decisions and uncertainties.

Level 5: MOP. MOP can lead to the Pareto set by solving repeatedly different optimization problems with new L_k values. The minimizing single objective optimization programming (MINSOOP) algorithm developed by Fu and Diwekar[11] also implements the HSS technique to generate combinations of the right-hand side L_k, and the MINSOOP algorithm is found to be superior to the existing algorithm in terms of efficiency and accuracy. This uniformity property of the HSS technique also contributes to accurate MOP solutions.

In summary, the computational efficiency of this MOP framework is improved at various levels by using the HSS, HSTA, and MINSOOP methods. All of the improvements exploit the uniformity and fast convergence properties of the HSS technique. As the first step in this framework, the HSTA algorithm in the MOP framework is used for designing candidate EBSs for HOAc extraction under uncertainty in the next section. Then the EBS selection is coupled with the IPS recycling problem and optimized simultaneously.

3. EBS Selection under Uncertainty

This section describes how to design candidate EBSs using the HSTA algorithm and presents important stochastic CAMD results. Because HOAc in this case study is a valuable solvent and can also be a pollutant when released to the environment, it is desirable to minimize the discharge of HOAc to the environment. To recycle HOAc to the upstream processes from the waste solvent streams, an extraction process is performed as shown in Figure 1. For the extraction process, one can use either high boiling point extracting agents[5,19,20] or low boiling point extracting agents,[2] depending on the process and facility being considered. Here, we focus on low boiling point extracting agents for HOAc because the current extracting agent is EtOAc, a low boiler.

Table 1. Uncertainty Analysis on the Estimation of Infinite Dilution Activity Coefficient by UNIFAC[22,23]

category of γ_{ij}^{∞}	distribution	mean	standard deviation	no. of data sets used
organic−water	log-normal	2.92	5.94	227
water−organic	normal	1.08	0.37	41
organic−organic	log-normal	1.14	1.30	161

3.1. Stochastic EBS Selection Model. To replace the current EBS molecules or to design a new one, there are several criteria to be considered, such as (a) distribution coefficient (m), (b) solvent selectivity (β), (c) solvent loss (S_L), (d) physical properties, (e) toxicology data such as LC_{50} and LD_{50}, (f) environmental properties such as persistence, bioconcentration factor (BCF), and reactivity, and (g) cost.[20] For extraction processes, the final selection of solvents will generally be dominated by m and β. Detailed definitions of m, β, S_L, and boiling point are described in Appendix A, and the first three solvent properties are functions of the infinite dilution activity coefficient (γ^{∞}) and molecular weight.

To predict m, β, and S_L, the (original) UNIFAC equation with recently revised parameters[21] is used. The interaction parameters (a_{kl}) between groups k and l in the UNIFAC equation are obtained from regressions of experimental data and are thus subject to uncertainty due to experimental and regression errors. Furthermore, the activity coefficient (γ) at a finite condition is, by definition, extrapolated to an infinite dilution activity coefficient (γ^{∞}) in which large discrepancies between experimental and calculated values may be observed.

It is not a good idea to define uncertainties in every a_{kl} parameter because a_{kl} lies in the deepest part of the nonlinear UNIFAC equation and there are too many a_{kl} parameters. (The number of a_{kl} in the UNIFAC equation[21] is 1803.) Instead, we defined an uncertainty factor (UF), the ratio of the experimental γ^{∞} to the calculated γ^{∞}, to handle uncertainties in the EBS selection model. We also divided γ^{∞} into three categories: organic−water, water−organic, and organic−organic. The subscripts in γ_{ij} mean the infinite dilution activity coefficient of component i in the component j phase. This division considers the difference between properties of water and those of organic chemicals. Table 1 summarizes uncertainty distribution types, means, and variances of the UF in these three categories and the number of data sets[22,23] used in this analysis. Figure 5 shows distributions of these UFs. We can see that there are large uncertainties in γ^{∞} predictions, especially in the $\gamma_{\text{organic,water}}^{\infty}$ prediction. Further, the mean values of UF greater than 1 show that the UNIFAC equation underestimates the infinite dilution activity coefficients. These uncertainty distributions are implemented in the combinatorial optimization problem under uncertainty,

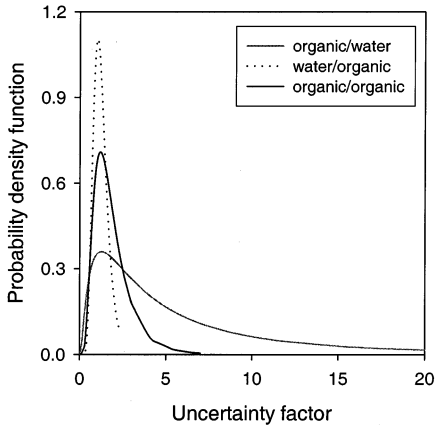

Figure 5. Probability density functions of UFs for the organic−water, water−organic, and organic−organic families.

which is formulated as follows:

$$\min -\frac{1}{N_{\text{samp}}}\sum_{j=1}^{N_{\text{samp}}}\left[\frac{\xi_j^1\,\gamma_{\text{B,A}}^{\infty}}{\xi_j^3\,\gamma_{\text{B,S}}^{\infty}}\right]\frac{\text{MW}_\text{A}}{\text{MW}_\text{S}} \tag{3}$$

subject to

$$\beta = \frac{1}{N_{\text{samp}}}\sum_{j=1}^{N_{\text{samp}}}\left[\frac{\xi_j^2\,\gamma_{\text{A,S}}^{\infty}}{\xi_j^3\,\gamma_{\text{B,S}}^{\infty}}\right]\frac{\text{MW}_\text{A}}{\text{MW}_\text{S}} \geq 7.0$$

$$S_\text{L} = \frac{1}{N_{\text{samp}}}\sum_{j=1}^{N_{\text{samp}}}\left[\frac{1}{\xi_j^1\,\gamma_{\text{S,A}}^{\infty}}\right]\frac{\text{MW}_\text{A}}{\text{MW}_\text{S}} \leq 0.058$$

$$47 \leq T_{\text{BP}}\,(°\text{C}) \leq 118$$

$$2 \leq N_1 \leq 10$$

$$1 \leq N_2^{(i)} \leq 24, \quad \forall\, i \in N_1$$

$$\xi^1 \sim \log N(2.92, 5.94)$$

$$\xi^2 \sim N(1.08, 0.37)$$

$$\xi^3 \sim \log N(1.42, 1.14)$$

where ξ is an uncertain parameter of the UF and is imposed on the estimated γ^{∞}. The constraint bounds of β and S_L are based on the values of the current extracting agent, EtOAc.

Discrete decision variables are the number of groups (N_1) in a solvent molecule and the group index ($N_2^{(i)}$) of that molecule, and by combining these decision variables, we can build a unique solvent molecule. By definition, solvent selection problems are to find the optimal values of N_1 and N_2 that correspond to the best solvent molecules. The set of groups (i.e., group indices) in Table 2 is specially designed for linear or branched hydrocarbons while aromatic, cyclic, and halogenated compounds are eliminated because of environmental concerns.

For molecular connectivity, a simple octet rule is applied. If nd_i represents the number of connecting nodes of a group index i, then the octet rule is

$$\sum_{i}^{N_1} \text{nd}_i = 2(N_1 - 1) \tag{4}$$

In addition, the maximum number of nd_i and the number of functional groups in a molecule can be restricted for faster and more practical molecular combinations.

Table 2. Set of Discrete Decision Variables for UNIFAC Group Contribution Method

i	$N_2^{(i)}$	i	$N_2^{(i)}$	i	$N_2^{(i)}$	i	$N_2^{(i)}$
1	CH_3-	7	$CH_2{=}C<$	13	CH_3CO-	19	CH_3O-
2	$-CH_2-$	8	$-CH{=}C<$	14	$-CH_2CO-$	20	$-CH_2O-$
3	$-CH<$	9	$>C{=}C<$	15	$-CHO$	21	$>CH-O-$
4	$>C<$	10	$-OH$	16	CH_3COO-	22	$-COOH$
5	$CH_2{=}CH-$	11	CH_3OH	17	$-CH_2COO-$	23	$HCOOH$
6	$-CH{=}CH-$	12	H_2O	18	$HCOO-$	24	$-COO-$

Under given chemical connectivity constraints, there are three processes to build a new group combination from the current combination: addition, contraction, and random bump. In the addition process ($N_1 = N_1 + 1$), the number of groups (N_1) in a solvent molecule is increased, and a random group index is assigned to that increased group. In the contraction process ($N_1 = N_1 - 1$), one group is randomly deleted. In the random bump process ($N_1 = N_1$), the number of groups in a molecule is unchanged. Instead, an arbitrarily selected group index ($N_2^{(i)}$) is randomly bumped up or down. The magnitude of these bumps is also random. The probabilities for these three processes are specified at 30%, 30%, and 40%, respectively. High random bump probability is assigned to guarantee the sampling of all of the group indexes.

3.2. EBS Selection Results. Because of the large combinatorial space of groups and large UFs in the infinite dilution activity coefficients, this problem represents a complex combinatorial optimization problem under uncertainty and a challenge to existing optimization techniques. This problem can be efficiently solved by the HSTA algorithm, the second level in the MOP framework (see Figure 3). For details, refer to the paper by the authors.[10]

Table 3 shows the optimal candidate EBSs from the deterministic and stochastic EBS selection models. From 40 candidate solvents generated, the first 15 solvents are ranked with respect to the order of m, and only eight solvents appear in both cases (see boldfaced solvents in this table). This implies that the deterministic case does not generate several or many promising solvents, which appeared in the stochastic case. Note that the current extracting solvent, EtOAc, is not listed in the top 15 solvents in both cases because it has a small m. At the deterministic case, EtOAc is outside the top 40 candidate solvents and at the stochastic case EtOAc is ranked as the 26th extracting agent (after rerunning of the deterministic model with 50 candidate solvents, it is listed at 46th with an m of 0.3156). As expected, distribution coefficients for the stochastic case are greater than those of the deterministic case because of the positively skewed UFs, mainly that of $\gamma_{\text{organic,water}}^{\infty}$.

The probability distribution function (pdf) analysis shows that the pdf of the deterministic case looks like a narrow log-normal distribution with a mean of 0.53 and a small variance, while the one of the stochastic

Table 3. Top 15 Candidate Solvents for the Deterministic and Stochastic Cases

	deterministic case		stochastic case	
no.	optimal solvents	*m*	optimal solvents	*m*
1	2**CH$_3$**, **CH$_2$**, **CH**, **HCOO**	0.87	2**CH$_3$**, **CH$_2$**, **CH**, **HCOO**	2.95
2	CH_3, $3CH_2$, HCOO	0.87	**CH$_3$**, **CH$_2$**, **CH=CH**, **HCOO**	2.60
3	**CH$_3$**, **CH$_2$**, **CH=CH**, **HCOO**	0.76	**CH$_3$**, **CH$_2$**, **CH$_2$=C**, **HCOO**	2.55
4	**CH$_3$**, 3**CH$_2$**, **CH$_2$=C**, **HCOO**	0.75	CH_3, CH_2=CH, $2CH_2O$	2.27
5	$2CH_2$, CH_2=CH, HCOO	0.72	**CH$_3$**, **CH$_2$=CH**, **CH$_3$O**, **CH−O**	2.27
6	CH_3, CH, CH_2=CH, HCOO	0.72	CH_3, CH, CH_2=CH, CH_3O, CH_2O	2.15
7	**CH$_3$**, **CH$_2$=CH**, **CH$_3$O**, **CH−O**	0.66	**CH$_3$**, **CH$_2$=C**, **CH$_3$O**	2.09
8	CH_3, CH, CH_2=CH, CH_2O, CH_3O	0.63	**CH$_3$**, **CH=CH**, **CH$_3$CO**	2.08
9	2**CH$_2$**, **CH$_2$=CH**, **CH$_2$O**, **CH$_3$O**	0.63	CH_3, CH_2, CH_2=C, CH_3O, CH_2O	2.04
10	**CH$_3$**, **CH$_2$=C**, **CH$_3$CO**	0.61	**CH$_3$**, **CH$_2$**, **CH$_3$CO**	1.96
11	**CH$_3$**, **CH=CH**, **CH$_3$CO**	0.61	CH_3, CH_2, CH_2=C, CH_3O	1.84
12	CH_2, CH_2=CH, CH_3CO	0.60	**2CH$_2$**, **CH$_2$=CH**, **CH$_2$O**, **CH$_3$O**	1.55
13	CH_3, CH, CH_2=CH, CH_3O	0.58	CH_3, CH=CH, CHO	1.51
14	$2CH_2$, CH_2=CH, CH_3O	0.58	CH_3, CH_2, CH, CH_2=CH, CH_3O	1.49
15	**CH$_3$**, 2**CH$_2$**, **CH$_3$**, **CO**	0.57	$2CH_3$, CH, CH_2=C, CH_3O	1.41

case has a wide normal distribution with a mean of 1.34 because of the positively skewed UFs. We can also find different types of solvent molecules in the stochastic case. Although most of the candidate solvents are formates, esters, and ethers, the stochastic case can also generate aldehydes, alkanes, and alkenes. New types of solvent molecules mean that the stochastic model can cover a wider range of combinatorial space of discrete decision variables (i.e., N_2).

The value of stochastic solution (VSS)[24] can be used to quantify the effects of uncertainty and represents the loss by not considering the uncertainty effects. VSS is the difference between taking the average value of the uncertain variable as the solution as compared to using stochastic analysis and optimizing the expected value. Because the UFs, represented by ξ over γ^∞, are implemented in the objective function, we can assume the expected value of the stochastic problem with the average ξ to be 2.06 times the deterministic m, as shown in Table 3. For the first set of solvent molecules in this table, the VSS of this case study is estimated as 1.16, and the stochastic optimization therefore increases the performance (distribution coefficient in this study) by 65%. Other sets of solvent molecules have similar VSS values. Thus, it makes sense to consider the stochastic solution for real implementations. Here, the solvents are used in the coupled EBS selection and IPS recycling framework described earlier.

4. Integrated EBS Selection and IPS Recycling Problem: An MOP Problem

4.1. Problem Formulation. The objectives of this MOP problem under uncertainty are to maximize HOAc recovery (Z_1), minimize EIs based on LC_{50} (Z_2) and based on LD_{50} (Z_3) (it should be remembered that there are numbers of other EIs like global warming, ozone depletion, photochemical oxidant formation, etc.; however, because we are looking for new solvents built from groups, we need group contribution methods for predicting other EIs. Unfortunately, in most cases group contribution methods for these impacts are not available. This is the reason we consider LC_{50} and LD_{50}, in which group contribution methods are available), and maximize the process flexibility (Z_4). The mathematical formulation of this MOP problem for the EBS selection and IPS recycling case study is

$$\min Z_1 = -\frac{\text{HOAc in product}}{\text{HOAc in feed}} = f(x,y) \qquad (5)$$

subject to

$$1 \le y_1 \le 7$$

$$3 \le y_2 \le 9$$

$$0.5 \le x_1 \le 0.95$$

$$20 \le x_2 \le 35$$

$$Z_2 = \text{EI} = \frac{\sum F_{\text{solvent,out}}}{\text{LC}_{50,\text{solvent}}} + \frac{\sum F_{\text{HOAc,waste}}}{\text{LC}_{50,\text{HOAc}}} \le L_2$$

$$Z_3 = \text{EI} = \frac{\sum F_{\text{solvent,out}}}{\text{LD}_{50,\text{solvent}}} + \frac{\sum F_{\text{HOAc,waste}}}{\text{LD}_{50,\text{HOAc}}} \le L_3$$

$$Z_4 = -\frac{\text{feasible runs}}{\text{total runs}} \le L_4$$

$$y_i = \text{integer}$$

where x and y are continuous and discrete decision variables. The continuous decision variable vector **x** is [split fraction, distillation bottom rate, EBS makeup flow rate, heat duty]^T, and the discrete decision design vector **y** is [solvent type, distillation feed point, number of equilibrium stages]^T. Although the current HSTA algorithm generated candidate solvents, as shown in Table 3, a small number of candidate EBSs are present in the Aspen Plus databank. By comparing the candidate EBSs with the Aspen Plus databank and by checking the presence of a binary azeotrope between water and EBS, we can find that seven candidate EBSs are in the databank for this process. Existence of a binary *heterogeneous* azeotrope between water and EBS is critical because the current process (see Figure 1) is based on the heterogeneous azeotropic distillation whose top azeotropic stream is separated in the decanter. The candidate solvents are (1) methyl propyl ketone, (2) methyl isopropyl ketone, (3) diethyl ketone, (4) EtOAc, (5) methyl propionate, (6) isopropyl acetate, and (7) propyl acetate, and their m, β, S_L, and boiling points are summarized in Table 4. Now, the HSTA algorithm is modified to generate only these solvents in order to

Table 4. Selection Criteria for Candidate EBSs[a]

EBS	m	β	S_L	T_{BP} (°C)	rank I	rank II
methyl propyl ketone	0.5740	30.86	0.053	102.3	15	10
methyl isopropyl ketone	0.5738	30.88	0.053	94.4	16	
diethyl ketone	0.5084	41.68	0.029	102.0	22	
ethyl acetate	0.3156	14.62	0.056	77.1		27
methyl propionate	0.2784	17.52	0.039	79.5		32
isopropyl acetate	0.2627	16.66	0.023	88.5		38
propyl acetate	0.2627	16.60	0.023	101.5		39

[a] Ranks I and II are obtained at the deterministic and stochastic EBS selection models, respectively.

eliminate redundant Aspen Plus simulations with a nondatabank component.

Discrete decisions include the total number of stages of the extractor and the distillation column (20 and 25, respectively), solvent type (y_1), and the feed point to the distillation column (y_2). Continuous decisions include the distillation bottom flow, split ratio (i.e., EBS recycle ratio), EBS makeup flow rate, and heat duties of the distillation column. The feed rate is assumed to be 100 kmol/h with the molar feed composition of 0.7/0.3 in water−HOAc, and process uncertainty is imposed on the feed flow rates.

LC_{50} and LD_{50} data (in Table 5) are needed to evaluate the EIs of the HOAc recovery process. Most of the LC_{50} and LD_{50} data are taken from these references,[25,26] while those whose values cannot be found were estimated using a new group contribution method. A detailed description of this method is given in Appendix B.

There is another important factor that should be considered in EBS selection and IPS recycling. In industrial practice, heat duty is critical in deciding alternatives for continuous distillation. Thus, before the MOP problem is solved, the proposed EBSs are also checked through the reboiler heat duty comparison in order to eliminate candidate EBSs that require high heat duty. From Figure 6, it can be seen that all extracting entrainers do not have large positive deviations from the heat duty of EtOAc, the current extracting agent, and the proposed EBSs can hence be used in the MOP problem.

4.2. Pareto Optimal Solution. The first step in solving MOP problems is to obtain a payoff table, which shows a potential range of values of each objective. A payoff table contains individual objective values (Z_k^*) for single optimization problems (k) and also provides potential ranges of the objectives on the Pareto surface (i.e., Z_L to Z_U). The minimum value (Z_L) of the Pareto surface is equal to the individual optimal value (Z_k^*), while the maximum value (Z_U) of the Pareto surface is the maximum value of that objective found when minimizing the other objectives individually. In this way, an approximated range of the right-hand side L_k in the Pareto surface is determined. Table 6 shows a payoff table for optimization under uncertainty, in which the variation in the feed flow rate is ±5% of the nominal feed flow rate and is normally distributed. To

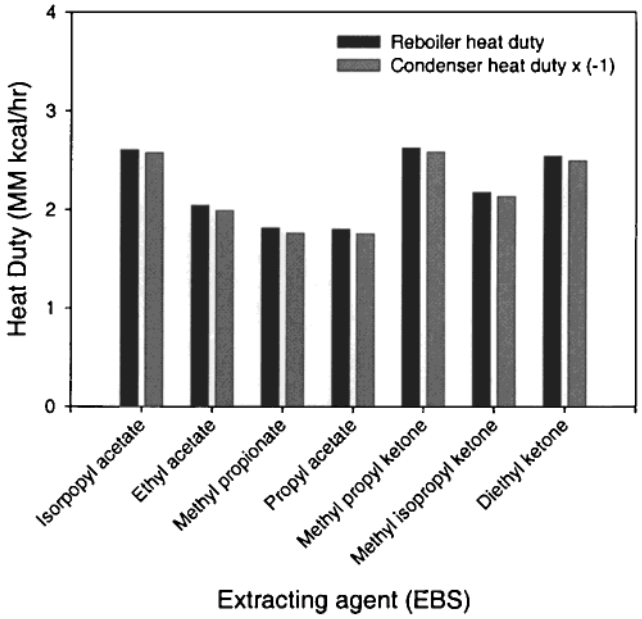

Figure 6. Reboiler and condenser heat duties of the candidate extracting agents.

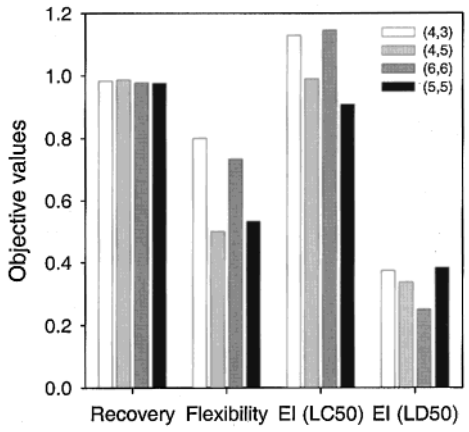

Figure 7. Pareto optimal solutions at 5% feed flow variations (numbers in the parentheses represent EBS type and distillation feed point).

clarify, the flow rate of each feed component is independently varied so that the total flow rate and its composition are altered.

Figure 7 shows the Pareto optimal solution of the MOP framework for the EBS selection and IPS recycling problem. Because of a highly discretized solution surface and large infeasible regions (in terms of yield and feasibility), only four Pareto solutions from 30 sets of optimization problems can be obtained. The first three EBSs in Table 4 are not in the Pareto optimal solutions because the azeotropic boiling points of these molecules with water are relatively high. EtOAc (4), which is the current extracting agent, has advantages in terms of two objectives: HOAc recovery (Z_1) and process flexibility (Z_4). High HOAc recovery can be predicted because the distribution coefficient of EtOAc is the

Table 5. LC_{50} and LD_{50} Data of Seven Candidate EBSs and HOAc

	methyl propyl ketone	methyl isopropyl ketone	diethyl ketone	ethyl acetate	methyl propionate	isopropyl acetate	propyl acetate	acetic acid
LC_{50} [mg/L]	1532	861	1540	230	922[a]	94[a]	60	88
LD_{50} [mg/kg]	3730	148	2140	5620	5000	3000	9370	3310

[a] Values estimated by the equation in Appendix B.

Table 6. Payoff Table at 5% Feed Flow Variation

k	objective	Z_k^*	Z_L	Z_U
1	HOAc recovery	−0.9869		
2	EI based on LC_{50}	0.9078	0.9078	1.4470
3	EI based on LD_{50}	0.2500	0.2500	0.3840
4	flexibility	−0.8000	−0.8000	−0.4667

Table 7. Pareto Optimal Solutions at Different Feed Flow Variations and Uncertainty Distributions

flow variation	normal distribution	uniform distribution
5%	4 Pareto solutions	6 Pareto solutions
	2 ethyl acetates	2 ethyl acetates
	isopropyl acetate	2 isopropyl acetates
	methyl propionate	2 methyl propionates
10%	2 Pareto solutions	4 Pareto solutions
	2 isopropyl acetates	2 isopropyl acetates
		2 methyl propionates

highest among the EBSs present in the Pareto optimal solutions. Hence, the popular use of EtOAc can be understood from these two objectives. For the EI based on LC_{50} (Z_2), methyl propionate (5) is the best solvent for this integrated problem. This is mainly due to the relatively high LC_{50} value and the lowest mole fraction in the effluent streams (i.e., the raffinate from the extractor and the aqueous stream from the decanter). The mole fraction of methyl propionate is $O(-3)$, while those of other EBSs are $O(-2)$. For the EI based on LD_{50} (Z_3), isopropyl acetate (6) is the best choice even though it has the smallest LD_{50} value among LD_{50} solvents. It is found that isopropyl acetate has the lowest fresh solvent makeup flow rate (e.g., ~ 0.36 kmol/h vs 1.27 kmol/h of EtOAc), and this is the main reason for the lowest Z_3 value. The process flexibility objective is highly dependent on the distillation feed point as well as EBS molecules. The second and third bars of process flexibility in Figure 7 show flexibilities when using EtOAc as an extracting agent. It is seen that a small change of the distillation feed point can significantly change the process flexibility. This phenomenon is also observed at different ranges of feed flow variations, which is described in the next paragraph.

If the feed flow variation is increased to 10%, then significantly different results in the Pareto optimal solution set can be obtained. Table 7 shows the MOP results at different feed flow variations and uncertainty distributions. The most distinctive feature of the result is that EtOAc is no longer in the Pareto optimal solution set. Only isopropyl acetate and methyl propionate are in the solution set. The second feature from the results is that the number of Pareto solutions is decreased. This means that decision makers (e.g., process engineers) have fewer choices. This reduced Pareto optimal set is mainly due to the large uncertainties in the feed flow rates, which can restrict the feasible region by increasing the ranges of L_k on the Pareto surface.

From this case study, we can see that the MOP framework for the simultaneous integration of EBS selection and IPS recycling resulted in better economic performance and environmental quality by improving EI (i.e., reducing solvent loss and using a safer solvent) and HOAc recovery without changing the process flowsheet structure. Thus, the proposed framework can provide different chemical and design alternatives to decision makers, and uncertainties in the framework can play a significant role in the early design and synthesis stages.

5. Conclusions

A simultaneous integration of EBS selection and IPS recycling is presented in terms of a multiobjective optimization framework for superior economic and environmental performances. In the EBS selection level, the HSTA algorithm is developed for designing solvent molecules under uncertainty. This algorithm can efficiently solve a combinatorial optimization problem under uncertainty and provide a different set of candidate solvents as compared to that of the deterministic solvent selection model. In the IPS recycling model, the HOAc recovery process is modeled in Aspen Plus with a nonlinear programming optimizer. Finally, the EBS selection and IPS recycling models are integrated under the MOP framework where an efficiently improved MOP algorithm is applied. The MOP framework can provide very distinctive chemical and design alternatives (i.e., Pareto set), and variations in the feed flow rate can greatly affect the size and shape of the Pareto surface. At small uncertainties in feed flow and composition, EtOAc, isopropyl acetate, and methyl propionate with optimal design and operating conditions can form the Pareto optimal solution set, but at large uncertainties, only isopropyl acetate and methyl propionate can form the Pareto set with fewer choices to decision makers. From this study, we can see the importance of simultaneous integration of EBS selection and IPS recycling under the multiobjective optimization framework, and this novel MOP framework can be applied to other large-scale stochastic mixed-integer nonlinear optimization problems because the two algorithms are computationally efficient even in the case of combinatorial explosion and uncertainty inclusion.

Acknowledgment

This work is supported by the NSF Goali Project (CTS-9729074). We also thank Dr. J. Siirola of Eastman Chemical Company, Kingsport, TN, for providing the case study.

Appendix A: EBS Selection Criteria

The distribution coefficient (m), a measure of the solvent capacity, is the most important factor and represents the solute distribution between being in the solvent and in the raffinate phases, as shown in the following equation:

$$m = \frac{x_{B,S}}{x_{B,A}} \frac{MW_A}{MW_S} \simeq \frac{\gamma^{\infty}_{B,A}}{\gamma^{\infty}_{B,S}} \frac{MW_A}{MW_S}$$

where γ^{∞} is the infinite dilution activity coefficient and MW is molecular weight. The symbols A, B, and S represent the raffinate, solute, and solvent phases, respectively. A high value of m reduces the size of the extracting equipment and the amount of recycling solvent.

Solvent selectivity (β), equivalent to the relative volatility in the distillation process, is the ratio between the distribution coefficients of solute and raffinate and is defined by

$$\beta = \frac{m_B}{m_A} \simeq \frac{\gamma^{\infty}_{A,S}}{\gamma^{\infty}_{B,S}} \frac{MW_B}{MW_A}$$

Table 8. Group Contribution Parameters (a, b_i) for LC_{50} Estimation

group	value	group	value	group	value
CH_3	−0.670 80	CH_3CO	−0.175 29	COO	0.000 00
CH_2	−0.534 20	CH_2CO	0.145 53	CH_3O	−0.430 48
CH	−0.233 05	CH−CO	0.000 00	CH_2O	0.429 43
C	0.337 05	CHO	−2.291 88	CH−O−	−0.113 61
CH_2=CH	−3.192 50	CH_3COO	−1.447 96	−COOH	0.000 00
OH	0.616 32	CH_2COO	−0.692 09	log K_{ow}	−0.013 06

Solvent selectivity estimates the ability of the solvent to selectively dissolve a solute, and hence a high β value reduces the cost of solute recovery.

Another important criterion is solvent loss (S_L), which is expressed by the following equation:

$$S_L = \frac{1}{\gamma^{\infty}_{S,A}} \frac{MW_S}{MW_A}$$

Low S_L means a high selectivity toward the solute and determines the immiscibility between solvent and raffinate. Among physical properties, the boiling point is a very important criterion too. This enables the use of similar energy consumption in the process and hence similar equipment size. Joback's method[27] is used to estimate the boiling points of candidate solvents:

$$T_{BP} = \sum_{i=1}^{N_1} t_a(N_2^{(i)}) + t_b \,(\text{unit: °C})$$

where t_a and t_b are temperature parameters. Two discrete decision variables, N_1 and N_2, represent the number of groups and group indices of a solvent molecule, respectively.

Appendix B: Prediction of LC_{50}

For solvents without experimental LC_{50} values, we developed a group contribution method based on the same UNIFAC groups and log K_{ow} (octanol−water partition coefficient) because the prediction accuracies of the existing quantitative structure−activity relationship models and other group contribution methods[28] are not highly accurate and/or require too many parameters to estimate the values.

Fathead minnow LC_{50} in mol/L for acyclic esters, ethers, ketones, aldehydes, and alcohols can be estimated by the following equation:

$$\log LC_{50} = a \log K_{ow} + \sum_i n_i b_i \quad (r^2 = 0.94;\ \text{data} = 46)$$

where a and b_i are parameters for log K_{ow} and group index i, respectively, and n_i is the number of occurrences of the group index i. Table 8 summarizes the parameters a and b_i, which are obtained from the regression of 46 experimental data. Strong contributions to aquatic toxicity come from terminal alkene, aldehyde, and acetate groups. Gao et al.[28] also found that aldehyde groups have the strongest toxicity contribution among the groups.

Table 9 shows prediction errors of several selected solvents. The average error in log LC_{50} prediction is 6.5%, and the maximum error is 39%. High prediction errors are mainly observed at high molecular weight molecules. This is one of the difficulties in group contribution methods for environmental properties.

Table 9. Comparison of Experimental and Estimated LC_{50} Values

chemical	MW	log K_{ow}	LC_{50} [mg/L]	log LC_{50} [mol/L] expt	log LC_{50} [mol/L] estd	error [%]
acetone	58	−0.24	8193	−0.8500	−0.8430	0.83
3-pentanone	86	0.99	1540	−1.7476	−1.7432	0.25
3,3-dimethyl-2-butanone	100	1.2	87	−3.0612	−1.8663	39.0
5-nonanone	142	2.71	31	−3.6617	−3.9024	6.58
butyl acetate	116	1.78	18	−3.8092	−3.7446	1.70
tert-butyl acetate	116	1.76	327	−2.5505	−3.1463	23.36
ethyl hexanonate	144	2.83	8.9	−4.2096	−4.2074	0.05
pentyl ether	158	4	3	−4.7025	−4.7038	0.03
ethanol	46	−0.24	14750	−0.4946	−0.5856	18.4
tert-butanol	74	0.35	6413	−1.0629	−1.0636	0.07
acetaldehyde	44	−0.34	34.3	−3.1086	−2.9582	4.84
hexyl acrylate	156	3.19	1.09	−5.1564	−6.5759	−27.53

For more accurate prediction of environmental properties, error analysis is required and mandates consideration on uncertainty in environmental property prediction as we have done in section 3, and this is one of the future works.

Literature Cited

(1) Downs, J. J.; Siirola, J. J. Challenges of Integrated Process and Control Systems Design for Operability. *1997 AIChE Annual Meeting* **1997**, 189a.

(2) Joback, K. G.; Stephanopoulos, G. Designing Molecules Possessing Desired Physical Property Values. *Foundations of Computer-Aided Process Design*; Snowmass Village, CO, 1989; pp 363−387.

(3) Gani, R.; Nielsen, B.; Freemasons, A. A Group Contribution Approach to Computer-Aided Molecular Design. *AIChE J.* **1991**, *37*, 1318.

(4) Buxton, A.; Livingston, A. G.; Pistikopoulos, E. N. Optimal Design of Solvent Blends for Environmental Impact Minimization. *AIChE J.* **1999**, *45*, 817.

(5) Hostrup, M.; Harper, P. M.; Gani, R. Design of Environmentally Benign Processes: Integration of Solvent Design and Separation Process Synthesis. *Comput. Chem. Eng.* **1999**, *23*, 1395.

(6) Ciric, A. R.; Huchette, S. G. Multi-objective Optimization Approach to Sensitity Analysis: Waste Treatment Costs in Discrete Process Synthesis and Optimization Problems. *Ind. Eng. Chem. Res.* **1993**, *32*, 2636.

(7) Lim, Y. I.; Floquet, P.; Joulia, X. Efficient Implementation of the Normal Boundary Intersection (NBI) Method on Multiobjective Optimization Problems. *Ind. Eng. Chem. Res.* **2001**, *40*, 648.

(8) Fu, Y.; Diwekar, U. M. Process Design for Environment: A Multi-objective Framework Under Uncertainty. *Clean Prod. Process.* **2000**, *292*, 92.

(9) Bhaskar, V.; Gupta, S. K.; Ray, A. K. Applications of Multiobjective Optimization in Chemical Engineering. *Rev. Chem. Eng.* **2000**, *16*, 1.

(10) Kim, K.-J.; Diwekar, U. M. Efficient Combinatorial Optimization Under Uncertainty−Part I. Algorithmic Development. *Ind. Eng. Chem. Res.* **2001**, accepted for publication.

(11) Fu, Y.; Diwekar, U. M. Efficient Multiobjective Optimization−Part I. Algorithm Development. *Oper. Res.* **2001**, submitted for publication.

(12) Cohon, J. L. *Multiobjective Programming and Planning*; Academic Press: New York, 1978.

(13) Chankong, V.; Haimes, Y. Y. *Multiobjective Decision Making Theory and Methodology*; Elsevier Science Publishing Co., Inc.: New York, 1983.

(14) AspenTech. *Aspen Plus Documentation*, Version 10.1-0; AspenTech: Cambridge, MA, 2000.

(15) Kalagnanam, J. R.; Diwekar, U. M. An Efficient Sampling Technique for Off-line Quality Control. *Technometrics* **1997**, *39*, 308.

(16) Kirkpatrick, S.; Gelatt, C. D., Jr.; Vecchi, M. P. Optimization by Simulated Annealing. *Science* **1983**, *220*, 671.

(17) Painton, L. A.; Diwekar, U. M. Stochastic Annealing for Synthesis Under Uncertainty. *Eur. J. Oper. Res.* **1995**, *83*, 489.

(18) Chaudhuri, P.; Diwekar, U. M. Process Synthesis Under Uncertainty: A Penalty Function Approach. *AIChE J.* **1996**, *42*, 742.

(19) Pretel, E. J.; López, P. A.; Bottini, S. B.; Brignole, E. A. Computer-Aided Molecular Design of Solvents for Separation processes. *AIChE J.* **1994**, *40*, 1349.

(20) Kim, K.-J.; Diwekar, U. M. Efficient Combinatorial Optimization Under Uncertainty—Part II. Application to Stochastic Solvent Selection. *Ind. Eng. Chem. Res.* **2001**, accepted for publication.

(21) Hansen, H. K.; Rasmussen, P.; Fredenslund, A.; Schiller, M.; Gmehling, J. Vapor−Liquid Equilibria by UNIFAC Group Contribution. 5. Revision and Extension. *Ind. Eng. Chem. Res.* **1991**, *30*, 2352.

(22) Tiegs, D.; Gmehling, J.; Medina, A.; Soares, M.; Bastos, J.; Alessi, P.; Kikic, I. *Activity Coefficients of Infinite Dilutions*; DECHEMA: Frankfurt, Germany, 1986; Vol. 6, Part 1.

(23) Gmehling, J.; Menke, J.; Schiller, M. *Activity Coefficients of Infinite Dilutions*; DECHEMA: Frankfurt, Germany, 1994; Vol. 6, Part 3.

(24) Diwekar, U. M. *Introduction to Applied Optimization*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2001; to be published.

(25) Agrawal, M. R.; Winder, C. Frequency and Occurrence of LD50 Values for Materials in the Workplace. *J. Appl. Toxicol.* **1996**, *16*, 407.

(26) Russom, C. L.; Bradbury, S. P.; Broderius, S. J.; Hammermeister, D. E.; Drummond, R. A. Predicting Modes of Action from Chemical Structure: Acute Toxicity in the Fathead Minnow (Pimephales promelas). *Environ. Toxicol. Chem.* **1997**, *16*, 948.

(27) Joback, K. G.; Reid, R. C. Estimation of Pure-Component Properties from Group-Contributions. *Chem. Eng. Commun.* **1987**, *57*, 233.

(28) Gao, C.; Govind, R.; Tabak, H. H. Application of the Group Contribution Method for Predicting the Toxicity of Organic Chemicals. *Environ. Toxicol. Chem.* **1992**, *11*, 631.

Received for review September 17, 2001
Revised manuscript received December 13, 2001
Accepted December 14, 2001

IE010777G