

Optimization Accounting for Property Prediction Uncertainty in Polymer Design

Costas D. Maranas

Department of Chemical Engineering
The Pennsylvania State University
University Park, PA 16802

Abstract- In this paper the problem of designing a polymer repeat unit with fine-tuned or optimized thermophysical and mechanical properties is addressed. The values of these properties are estimated using group contribution methods based on the number and type of the molecular groups participating in the polymer repeat unit. By exploiting the prevailing mathematical features of the structure-property relations the following two research objectives are addressed: (i) How to efficiently locate a ranked list of the best polymer repeat unit architectures with mathematical certainty; and (ii) how to quantify the effect of imprecision in property estimation in the optimal design of polymers. A blend of mixed-integer linear optimization, chance-constrained programming, and multilinear regression is utilized to answer these questions. The proposed methodology is highlighted with a illustrative example. Preliminary results (see also Maranas (1996a,b)) demonstrate that the proposed framework identifies the the mathematically best molecular design and quantifies the profound effect that property prediction uncertainty may have in optimal molecular design.

INTRODUCTION

Optimal computer-aided molecular design (CAMD) involves the identification of a single molecule or mixture of molecules that optimizes one or multiple objectives while satisfying a number of macroscopic property specifications. In polymer engineering it is often important to systematically identify the architecture of a polymer which best meets a number of performance requirements and design considerations. When there are only a few "loose" property constraints, a suitable candidate might be found from polymer property databases and tabulations. However, when many "tight" property constraints must be satisfied simultaneously or a performance objective needs to be maximized, database searching, which is limited by the number of tabulated alternatives, is not always sufficient. The advantage of CAMD is that it is not restricted to an existing tabulation of polymer designs. Instead, by utilizing structure-property prediction models such as group contribution methods, (van Krevelen, 1990; Joback and Reid 1987; Constantinou and Gani, 1994), to estimate physical, chemical and mechanical polymer properties, CAMD is capable of eliciting new, sometimes unexpected, molecular designs. This aids the preliminary screening process, which must be followed by experimental testing and verification with published property data. A number of computer-aided approaches have been proposed in the chemical engineering literature for the design of polymers. These include enumeration techniques (Joback, 1985; Derringer and Markham, 1985), genetic algorithms (Venkatasubramanian 1994), MINLP optimization (Vaidyanathan and El-Halwagi, 1996), MILP optimization for linear structure-property models (Constantinou *et al*, 1995), and MILP reformulations for certain classes of nonlinear structure-property models (Maranas 1996a).

Identifying the best molecular design with mathematical certainty is important because by eliminating the caveat of convergence to suboptimal molecular designs the chances of identifying novel, possibly counter-intuitive, superior design alternatives are improved. Fur-

thermore, by removing the possibility of unknowingly converging to suboptimal solutions, possible discrepancies between obtained optimal solutions and experimentally derived designs can be explicitly and unequivocally attributed to property estimation imprecision (uncertainty). The discrepancies between group contribution predictions and experimental data are typically 5–10% or even higher (van Krevelen, 1990). The frequency and magnitude of these discrepancies provide ample motivation for quantifying their effect on optimal molecular design. Without properly quantifying property prediction uncertainty, one can have only partial confidence in the obtained solutions.

The following two objectives are addressed herein:

1. Identify with mathematical certainty a ranked list of the best polymer repeat units architectures.
2. Given a probabilistic description of property prediction uncertainty, quantify its effect on the selection of the best polymer design.

OPTIMAL POLYMER DESIGN

The optimal polymer design problem is first considered assuming that the structure-property group contribution models are exact. While the proposed framework is generally applicable, initial efforts have focused on the polymer design problem. A detailed analysis can be found in (Maranas, 1996a). The essential features of the optimal polymer design problem can be captured mathematically in the following mixed-integer nonlinear optimization problem:

$$\begin{aligned} \min \quad & \mathcal{MP}(p_j(\mathbf{n})) && \text{(OMD)} \\ \text{subject to} \quad & p_j^L \leq p_j(\mathbf{n}) \leq p_j^U, \quad j = 1, \dots, M \\ & n_i \in \{n_i^L, n_i^L + 1, \dots, n_i^U\}, \quad i = 1, \dots, N \end{aligned}$$

In formulation **(OMD)**, $\mathbf{n} = (n_1, \dots, n_N)$ is the vector of the integer variables $n_i \in \{0, 1, 2, \dots\}$ describing the number of times the i^{th} molecular group participates in the molecule. The expressions $p_j = p_j(\mathbf{n})$, $j =$

1, ..., M, established by group contribution methods, denote the functionality between property j and the type and number of different molecular groups. The objective function \mathcal{MP} is a measure of performance of the molecular design and is typically a function of one or more properties, $\mathcal{MP} = \mathcal{MP}(p_j(\mathbf{n}))$. Additional constraints may be placed on (OMD) to ensure structural feasibility. The following two most widely used measures of performance are considered:

1. Minimization of the maximum allowable scaled deviation of properties from some target values (property matching formulation (PM)) which is appropriate for optimally designing a polymer with properties that match some prespecified targets:

$$\begin{aligned} \min \quad & s \\ \text{s.t.} \quad & s \geq \frac{1}{p_j^s} \left| \frac{\sum_{i=1}^N A_{ij} n_i}{\sum_{i=1}^N B_{ij} n_i} - p_j^o \right|, \quad j = 1, \dots, M \end{aligned}$$

Here s is the maximum scaled violation over all target properties, p_j^o is the target value for property j and p_j^s is an appropriate scale. Typically, the maximum percent property deviation is minimized, and therefore $p_j^s = p_j^o$.

2. Minimization/maximization of a single property j^* (property optimization (PO) formulation) which is applicable when the identification of a polymer with the largest or smallest value for one property (e.g., density) is sought while maintaining the other property values within lower and upper bounds.

$$\begin{aligned} \max / \min \quad & p_{j^*} = \frac{\sum_{i=1}^N A_{ij^*} n_i}{\sum_{i=1}^N B_{ij^*} n_i} \\ \text{s.t.} \quad & p_j^L \leq \frac{\sum_{i=1}^N A_{ij} n_i}{\sum_{i=1}^N B_{ij} n_i} \leq p_j^U, \\ & j = 1, \dots, M \end{aligned}$$

where variable p_{j^*} denotes the j^* th property to be minimized or maximized and p_j^L, p_j^U are the corresponding lower and upper bounds.

The group contribution methods compiled by van Krevelen (1990) and extensively used by other researchers (Venkatasubramanian *et al*, 1995; Vaidyanathan and El-Halwagi, 1995) are utilized in this study. In (Maranas, 1996a) it is shown that most of thermophysical, optical, electromagnetic and mechanical property estimating formulae for polymers conform or can be transformed to the following general functionality,

$$p_j(\mathbf{n}) = \left(\frac{\sum_{i=1}^N A_{ij} n_i}{\sum_{i=1}^N B_{ij} n_i} \right)^{d_j}, \quad j = 1, \dots, M. \quad (I)$$

Here A_{ij} and B_{ij} are given group contribution parameters associated with a specific molecular group i and property j . Formulation (OMD) is, in general, very difficult to solve due to nonlinearities in the property–structure relations and the large number of ways that a set of molecular groups can be interconnected in a structurally feasible manner (Joback and Stephanopoulos, 1989). A systematic analysis framework is introduced (see (Maranas,

1996a) for details), for transforming the original formulations (PM) and (PO) with nonlinear structure–property functionalities of type (I) into equivalent mixed–integer linear (MILP) problems. This is accomplished by (i) eliminating absolute values, (ii) expressing integer variables as the sum of binaries, and (iii) eliminating products of binary and continuous variables through their equivalent representation with linear constraints. The key advantage of reformulating (PM) and (PO) as MILP problems is that efficient MILP solvers (e.g., CPLEX) can be utilized to locate the global optimum molecular design with mathematical certainty. Moreover, by incorporating appropriate integer cuts in the formulation not only can the best solution be found, but the second best, third best, etc. can be generated successively. Next, a motivating example is addressed to highlight the proposed optimization framework.

Illustrative Example

This small–scale example involves the design of a polymer repeat unit which meets constraints on density, water absorption, and glass transition temperature (Derringer and Markham, 1985). The molecular groups which are allowed to participate in the polymer repeat unit are:

Index	1	2	3	4
Group	-CH ₂	-O-	-COO-	-O-
Index	5		6	7
Group	-CONH-		-CHOH-	-CHCl-

The contribution of these molecular groups to the three properties of interest (density, glass transition temperature and water absorption respectively), follow the empirical equations compiled by van Krevelen (1990):

$$\begin{aligned} D &= \left(\sum_{i=1}^N M_i n_i \right) \left(\sum_{i=1}^N V_i n_i \right)^{-1} \quad (g/cm^3) \\ T_g &= \left(\sum_{i=1}^N Y_{gi} n_i \right) \left(\sum_{i=1}^N M_i n_i \right)^{-1} \quad (K) \\ W &= \left(\sum_{i=1}^N 18H_i n_i \right) \left(\sum_{i=1}^N M_i n_i \right)^{-1} \\ &\quad (g H_2O/g polymer). \end{aligned}$$

The property targets for the (PM) formulation are,

$$\begin{aligned} W^o &= 0.005 \quad (g H_2O/g polymer), \\ D^o &= 1.50 \quad (g/cm^3), \\ T_g^o &= 110 \quad (K) \end{aligned}$$

The property scales W^s, D^s, T_g^s are selected to be equal to the property targets W^o, D^o, T_g^o . This ensures that per cent deviations from the target values are penalized equally for all properties. The property bounds for the (PO) formulation are:

$$0 \leq W \leq 0.18, \quad 298 \leq T_g \leq 673, \quad \text{and} \quad 1 \leq D \leq 1.5.$$

Both formulations are solved with the GAMS/CPLEX (Brooks *et al*, 1990) interface.

First the property matching problem (PM) is considered. Based on the analysis presented in (Maranas, 1996a), the original nonlinear, nonconvex formulation is transformed into an (MILP). Solution of the (MILP) formulation yields the following five best molecular designs shown here in decreasing order of optimality (CPU less than 1 second).

Table 1: The five best solutions for the (PM) problem

Alias	Repeating Unit	Violation
1-2	$-(\text{CH}_2-(\text{CHCl})_2)-$	0.0163
1-3	$-(\text{CH}_2-(\text{CHCl})_3)-$	0.0263
2-3	$-((\text{CH}_2)_2-(\text{CHCl})_3)-$	0.0526
0-1	$-(\text{CHCl})-$	0.1134
1-1	$-(\text{CH}_2-\text{CHCl})-$	0.1169

Note that all five best molecular designs involve only molecular groups $-\text{CH}_2-$ and $-\text{CHCl}-$ in 1-2, 1-3, 2-3, 0-1 and 1-1 ratios respectively. The predicted maximum scaled property violations range from about 2% for the best to 12% for the fifth best. Clearly the solution obtained depends on the adopted scaling; therefore it is very important to select the scaling property factors in a way that truly reflects their relative importance.

Next the property optimization problem (**PO**) is addressed involving the minimization of the water absorption subject to lower and upper bounds on glass transition temperature and density. The five best molecular designs in decreasing order of optimality are:

Table 2: The five best solutions for the (PO) problem

Alias	Repeating Unit	W
3-1	$-((\text{CH}_2)_3-\text{CHCl})-$	0.00318
2-1	$-((\text{CH}_2)_2-\text{CHCl})-$	0.00368
3-2	$-((\text{CH}_2)_3-(\text{CHCl})_2)-$	0.00401
1-1	$-(\text{CH}_2-\text{CHCl})-$	0.00441
2-3	$-((\text{CH}_2)_2-(\text{CHCl})_3)-$	0.00474

Note that all five best molecular designs involve again only molecular groups $-\text{CH}_2-$ and $-\text{CHCl}-$ in 3-1, 2-1, 3-2, 1-1 and 2-3 ratios respectively. The predicted minimum water absorptions range from 0.00318 to 0.00474 ($g \text{ H}_2\text{O}/g \text{ polymer}$).

Results on larger case-studies involving as many as 30 molecular groups can be found in (Maranas, 1996a). There it is observed that while in some cases the best design is clearly superior to the second, third, etc. in most cases such a clear cut distinction is absent. Despite the prevailing proximity in objective value the corresponding molecular architectures are often very different. This allows flexibility in experimental follow-up to consider cost, ease of synthesis, etc. which is not available should only the best molecular design be reported.

PROPERTY PREDICTION UNCERTAINTY

The work highlighted above removes the possibility of unknowingly converging to suboptimal solutions for the structure-property functionalities at hand. Therefore, any discrepancies between obtained solutions and experimental verifications can be attributed to property prediction uncertainty. This section describes how to assess the effect of property prediction uncertainty in optimal polymer design. Group contribution methods are based on the conjecture that molecular property values are uniquely defined by the molecular or other groups composing the molecule (transferability assumption). However, discrepancies between model prediction and actual experimental data imply that the transferability assumption is only approximately accurate because in reality there is always some disagreement between experimental measurements and group contribution predictions. These discrepancies can be described by recognizing that the contribution of molecular groups to various properties is not unchanged (independent of molecular architecture) but varies slightly around some nominal value depending on the particular molecular architecture. This intuitive abstraction can

be explicitly quantified by utilizing multivariate probability density distributions (possibly correlated) to model the likelihood of different realizations of the employed group contribution parameters A_{ij} and B_{ij} . The proposed probabilistic description of uncertainty renders both design specifications and performance objective stochastic, yielding the following optimal molecular design problem under uncertainty (**SOMD**):

$$\begin{aligned} & \max && \mathcal{MP}^{target} \\ \text{subject to} && \text{Prob} [\mathcal{MP} (p_j(\mathbf{n})) \geq \mathcal{MP}^{target}] \geq \alpha \\ && \text{Prob} [p_j^L \leq p_j(\mathbf{n}) \leq p_j^U] \geq \beta, \\ && j = 1, \dots, M \end{aligned}$$

Formulation (**SOMD**), which is much harder to solve than (**OMD**), involves a set of constraints imposing lower bounds on the probability of satisfying the performance objective and the imposed property lower and upper bounds. These constraints are called *chance-constraints*. (**SOMD**) identifies the maximum value of the performance target \mathcal{MP}^{target} which the stochastic performance objective \mathcal{MP} can meet with probability of at least α (e.g., 90%), and at the same time maintain all property values within their respective lower and upper bounds with probability greater than or equal to β . Therefore, the solution of (**SOMD**) will have at least an α chance of meeting the performance objective and at least a β chance of maintaining the property values within their designated bounds. By solving (**SOMD**) for different values of α and β , trade-offs between the performance objective target \mathcal{MP}^{target} , the probability α of meeting this performance target, and the probability β of satisfying all property constraints can readily be established.

Two special instances of (**SOMD**) are considered which extend the property matching (**PM**) and property optimization formulations (**PO**), discussed earlier, to account for property prediction uncertainty.

- (i) The stochastic property matching problem (**SPM**) which identifies the type and number of molecular groups for which a scaled property violation target s , guaranteed to be met by all properties j with probability of at least α , is minimized.
- (ii) The stochastic property optimization problem (**SPO**) which identifies the type and number of molecular groups for which a property target p_{j^*} for property j^* , met with probability α , is maximized while lower and upper bounds for the rest of the properties are satisfied with probability of at least β .

Formulations (**SPM**) and (**SPO**) involve probability terms whose evaluation for each realization of the deterministic variables requires the integration of multivariate probability density distributions. Many integration methods exist, but in general, they all exact a heavy computational burden either in the form of additional variables (quadrature point integration) or excessive function evaluation (Monte Carlo integration) and thus are restricted to problems with only a few uncertain parameters (less than ten). However, optimal polymer design problems routinely involve from tens to hundreds of uncertain parameters. To deal with such a high number of possibly correlated uncertain parameters the exact transformation of the original stochastic constraints into equivalent deterministic ones is accomplished based on the ideas pioneered by Charnes and

Cooper (1959). Assuming that the uncertain parameters A_{ij} and B_{ij} follow *stable* two-parameter probability density distributions (i.e., normal, Poisson, Chi-square, binomial, etc.), the stochastic (SOMD) can be transformed into an equivalent deterministic (MINLP) formulation with a linear binary and convex continuous part with separability (Maranas, 1996b). The resulting convex (MINLP) formulation can be solved to global optimality with existing decomposition-based algorithms such as OA/ER and GBD. This enables for the first time the exact solution of optimal molecular design problems under property prediction uncertainty. For the property matching problem (SPM), preliminary results indicate that property prediction uncertainty has a significant effect on the selection of the best molecular design. For example, frequent *reversals* of the order of optimality, found by the deterministic model, occur *and depend* on the selected probability α of satisfying the chance-constraints. This quantitatively demonstrates the intuitive expectation that the *answer to the question of what is the best molecular design depends on how often design targets are allowed to be violated*.

Next a systematic procedure is highlighted based on multilinear regression for calculating unbiased estimators (means) and sample estimators of the full variance-covariance matrix for the group contribution parameters.

Evaluation of the variance/covariance matrix

The means and variances/covariances of the group contribution parameters can be found by regressing separately the numerators and denominators of the adopted property prediction functionality. This simplifies the regression process by maintaining the linearity of the regression models without sacrificing accuracy.

Let $k = 1, \dots, K$ denote a set of molecular compounds (or repeat units) and \hat{p}_k the experimental measurement of property k for compound k . Assuming an additive linear group contribution relation between the times n_{ik} molecular group i participates in compound k we have,

$$p_k = \sum_{i=1}^N a_i n_{ik}, \quad k = 1, \dots, K$$

where p_k is the group contribution estimate of property p and a_i the group contribution parameters. (Multi)linear regression is based on the assumptions that for each specific X there is a normal distribution of Y which is (i) independent of X , (ii) whose mean depends linearly on X , and (iii) and with the same variance from which sample values of Y are drawn at random. Multilinear regression can be utilized to identify unbiased estimators (means) and sample estimators of the variance-covariance matrix of the vector of uncertain group contribution parameters $\mathbf{a} = [a_1, \dots, a_n]^T$. The minimization of the sum of the squares of the differences between the experimentally measured \hat{p}_k and estimated values p_k

$$\min \sum_{k=1}^K (\hat{p}_k - p_k)^2 = \sum_{k=1}^K \left(\hat{p}_k - \sum_{i=1}^N a_i n_{ik} \right)^2$$

yields the unbiased estimators for the values of the group contribution parameters (Snedecor and Cochran, 1989),

$$\mu(\mathbf{a}) = (\mathbf{N}^T \mathbf{N})^{-1} \mathbf{N}^T \mathbf{p}$$

where \mathbf{a} is the $(N \times 1)$ vector of the group contribution parameters,

$$\mathbf{a} = [a_1, a_2, \dots, a_N]^T,$$

\mathbf{N} is the $(K \times N)$ matrix whose elements are the integers n_{ik} counting how many times molecular group i participates in compound k , and \mathbf{p} is the $(K \times 1)$ vector of the experimental measurements,

$$\mathbf{p} = [p_1, p_2, \dots, p_K]^T.$$

Based on the analysis of Snedecor and Cochran (1989), the $(N \times N)$ variance-covariance matrix of the group contribution parameter vector \mathbf{a} is given by:

$$\text{Var}(\mathbf{a}) = (\mathbf{N}^T \mathbf{N})^{-1} s_{p,N}^2$$

where

$$s_{p,N}^2 = \frac{1}{K - N} \sum_{k=1}^K (\mathbf{p} - \mathbf{N}\mathbf{a})^T (\mathbf{p} - \mathbf{N}\mathbf{a})$$

is the unbiased estimator of the variance of the experimental property values. The identification of the vector of means and the variance-covariance matrix requires only the solution of linear systems of equalities, thus, very large volumes of data can be processed efficiently. Next the example considered earlier is revisited to highlight the proposed framework of analysis and support these observations.

Revisited Example

The same example addressed earlier is reconsidered after taking into account property prediction uncertainty. The group contribution parameters are assumed to be independent random variables (covariances equal to zero), normally distributed and with mean values equal to their reported values. Their variances are chosen to reflect the relative accuracy of the group contribution methods. For instance, density estimates based on group contribution are typically more accurate than estimates of water absorption or glass transition temperature. Specifically, the variance $Var(M_i)$ of M_i is chosen to be zero because the repeat unit molecular weight is rigorously additive to the individual molecular group contributions. The variance of V_i is selected so that 99% of possible realizations of V_i are within $\pm 5\%$ from the mean value $\mu(V_i)$, (see Kreyszig (1993)). Furthermore, a 10% scatter around the mean value for H_i and a 20% for Y_{gi} are imposed. Note that these values are not regressed based on experimental data but rather are selected arbitrarily to provide reasonable estimates for individual property uncertainties. The equivalent deterministic formulation of (SPM) for a set of different values for s_o ranging from 0.0163 (solution of (PM)) to 0.30 is constructed. Because the resulting (MINLP) involves a convex continuous and a linear discrete component which are mutually separable, the GAMS/DICOPT interface is guaranteed to identify the global minimum (single minimum). The results for different values of s_o are as follows:

Table 3: Results for the (SPM) problem

α	s_o	Repeat Unit	CPU s.
0.1761	0.0163	$-(\text{CH}_2-(\text{CHCl})_2)-$	32.72
0.2152	0.0200	"	21.31
0.4151	0.0400	"	21.31
0.5874	0.0600	"	9.27
0.7331	0.0800	$-(\text{CH}_2)_2-(\text{CHCl})_3-$	15.02
0.8349	0.1000	"	37.72
0.9043	0.1200	"	8.82
0.9481	0.1400	"	10.63
0.9737	0.1600	"	7.55
0.9876	0.1800	"	8.58
0.9945	0.2000	"	8.65
0.9999	0.3000	"	9.86

These results show that, as expected, the higher the probability α the larger (more conservative) the maximum scaled property violation target for the polymer repeat unit. The best polymer design for scaled property violations of less than about 0.065 (or α of less than 0.6) is the design identified as the best for the deterministic model, $-(\text{CH}_2-(\text{CHCl})_2)-$. However for probabilities of greater than 0.6 the best design becomes the third best design of the deterministic model $-(\text{CH}_2)_2-(\text{CHCl})_3-$. This result demonstrates that property prediction uncertainty may affect the selection of the best molecular design by reversing the deterministic order of optimality for certain probability levels α . In other words, *the answer to the question of what is the best molecular design depends on how often design targets are allowed to be violated*. Note that there is less than 18% chance of meeting the scaled property violation target 0.0163 predicted by the deterministic model. For a more likely target ($\alpha = 0.5$), a three-fold increase in the value of the scaled property violation target is predicted. Next, the trade-off curves between scaled property violation and probability are generated for the five best molecular designs predicted by the deterministic model (See Figure 1). Simple inspection of the trade-off curves reveals that the first three molecular designs with aliases 1-2, 1-3, 2-3 are superior to the ones with 0-1, 1-1 over the entire probability range. Additionally, while for probability less than about 0.6 design 1-2 is superior for higher probability values design 2-3 becomes the optimum. Note that this "crossing-over" of the trade-off curves of designs 1-2 and 2-3 is not unique. For example the curve for design 1-1 crosses over the one for design 0-1. The trade-off curves of all molecular designs start from the solution derived by the deterministic model reflecting that *the quantitative effect of property prediction uncertainty is to penalize the deterministic model predictions*. The probability of meeting the scaled property violation target postulated by the deterministic model is less than fifty per cent for all molecular designs. In fact the higher ranked the molecular design is, the lower the probability of meeting the deterministic model prediction appears to be.

Next the property optimization problem under uncertainty (SPO) is revisited involving the minimization of the water absorption subject to lower and upper bounds on glass transition temperature and density. First, the maximum value of β is identified for which the five best molecular designs 3-1, 2-1, 3-2, 1-1 and 2-3 satisfy the lower and upper bounds on density and glass transition temperature. The solution of the deterministic equivalent representation yields:

Table 4: Max β for the five best solutions

Rank	Repeating Unit	β_{max}
1	$-(\text{CH}_2)_3-\text{CHCl}-$	0.7505
2	$-(\text{CH}_2)_2-\text{CHCl}-$	0.9475
3	$-(\text{CH}_2)_3-(\text{CHCl})_2-$	0.9826
4	$-(\text{CH}_2-\text{CHCl})-$	0.9954
5	$-(\text{CH}_2)_2-(\text{CHCl})_3-$	0.9878

These values indicate that molecular designs 2-1, 3-2, 1-1 and 2-3 are more likely to satisfy the lower and upper bounds on density and glass transition than the best design according to the deterministic model 3-1. Thus, if for example a probability requirement on property bounds satisfaction of 0.90 is imposed then molecular design 3-1 must be excluded from consideration.

After fixing β for each one of the molecular designs 3-1, 2-1, 3-2, 1-1 and 2-3 to the values shown in Table 4, the stochastic property optimization formulation is solved (n_i 's fixed) while varying the probability level α between 0.1 to 0.9999. This yields the trade-off curves between minimum water absorption target and probability α of satisfaction of target for the five best molecular designs (See Figure 2). Clearly, the optimality order derived by the deterministic model is maintained and no "cross-over" points between trade-off curves are observed. The effect of uncertainty is not nearly as pronounced as in the property matching problem. In fact, a scatter of only about 0.0002 around the value for ($\alpha = 0.5$) of the water absorption target is observed for lower or higher values of α . For the group contribution parameter variances at hand, the results indicates how unlikely it is to have the optimality rankings derived by the deterministic model changed.

SUMMARY AND CONCLUSIONS

In this work a prototypical polymer problem with structure-property relations involving the ratio of two linear expressions in n_i raised to some real power is addressed. A reformulation procedure was described for recasting the original nonlinear problem into a (MILP) formulation which can be solved to global optimality with existing solvers. Next, the concept of utilizing probability density distribution function to model the uncertainty of group contribution parameters was introduced. The resulting chance-constraint formulation was then transformed into a deterministic equivalent (MINLP) problem with a convex continuous part. A small illustrating example was included to highlight the proposed framework. Results indicated the importance of not missing any optimal designs and the profound effect the property prediction uncertainty can play in optimal molecular design. The proposed methodology is not restricted to the specifics of the polymer design problem and efforts are currently underway to extend in other molecular design problems.

REFERENCES

- Brooke, A., Kendrick D. and Meeraus, A., 1988, GAMS: A User's Guide. Scientific Press, Palo Alto, CA
- Charnes, A. and Cooper, W.W., 1959, Management Science, **6**, 73
- Constantinou, L., Bagherpour, R., Gani, R., Klein, J. and Glen, R., 1996, Comput. chem. Engng. **20**, 685
- Constantinou, L. and Gani, R., 1994, AIChE J. **40**, 1697
- Derringer, G.C. and Markham, R.L., 1985, Journal of Applied Polymer Science **30**, 4609
- Joback K.G., 1989, *Designing Molecules Possessing Desired Physical Properties*. Ph.D. Thesis, MIT, Cambridge, MA.

Joback, K.G. and Reid, R.C., 1987, Chem. Eng. Commun. **57**, 233
 Joback, K. and Stephanopoulos, G., 1989, *FOCAPD '89*, 363, Snowmass, Co.
 Kreyszig, E., 1993, *Advanced Engineering Mathematics*, John Wiley & Sons, Inc., New York, NY, 7th edn.
 Maranas, C.D., 1996, Ind. Chem. Eng. Res. **35**, 3403
 Maranas, C.D., 1997, Optimal molecular design under property prediction uncertainty, accepted in *AIChE J.*
 Snedecor, G.W. and Cochran, W.G., 1989, *Statistical*

Methods. Iowa State University Press, Ames, Iowa, 8th edn.
 Vaidyanathan, R. and El-Halwagi, M., 1996, Ind. Eng. Chem. Res. **35**, 627
 van Krevelen D.W., 1990, *Properties of Polymers: their correlation with chemical structure; their numerical estimation and prediction from additive group contributions*. Elsevier, Amsterdam, 3rd edn.
 Venkatasubramanian, V., Chan, K. and Caruthers, J.M., 1994a, *Computers chem. Engng.* **18**, 833.

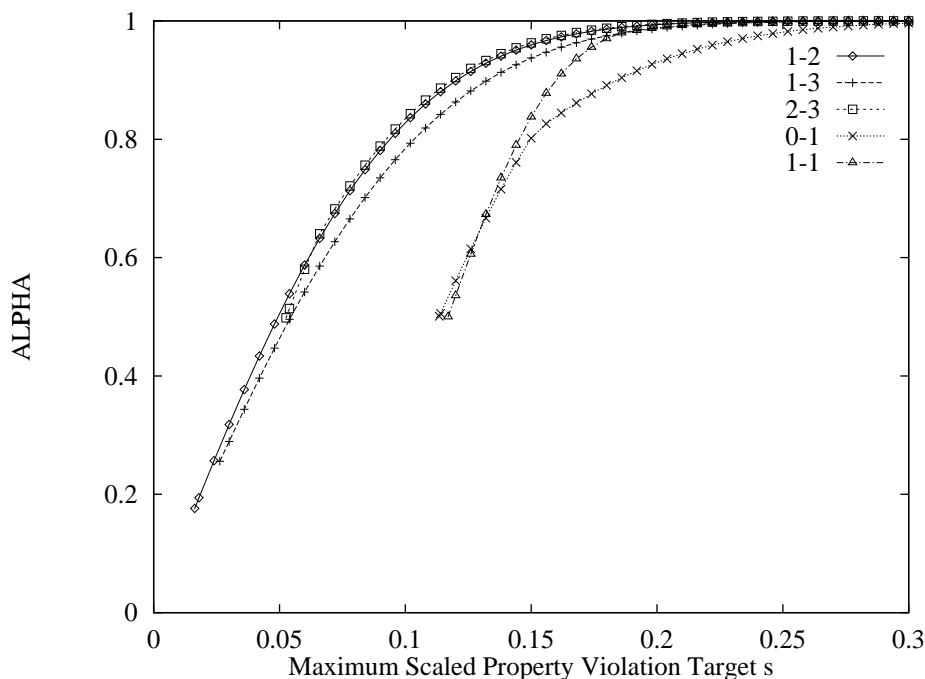


Figure 1: Trade-off curves of the five best molecular designs of (SPM).

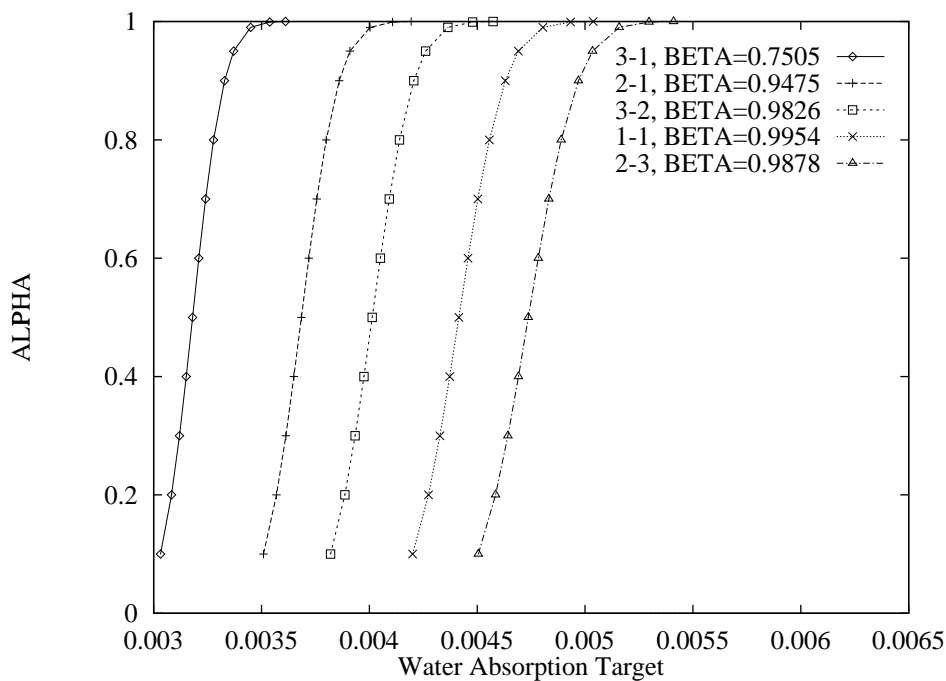


Figure 2: Trade-off curves of the five best molecular designs of (SPO).