Concept of System Similarity in Chemical Engineering

Fuzzy set theory is used to describe similarity of chemical engineering systems. The development of system similarity is given, and methods of determining a similarity relation are discussed. The system similarity relation is applied to the problems of measuring correlation confidence and quantifying modeling error. Sample problems are examined.

W. L. Kubic, Jr., F. P. Stein
Department of Chemical Engineering
Lehigh University
Bethlehem, PA 18015

Introduction

Chemical engineers are often confronted with immense quantities of information in the form of data, theories, and correlations. For example, compilations of physical property data contain the results of thousands of experimental studies. In spite of these vast amounts of information, data pertinent to the system of interest are often lacking. Knapp (1986) noted this problem when commenting on electronic data banks: "The information we have is not what we want. The information we want is not what we need. The information we need is not what we have."

Not having the necessary information does not mean that the available information is useless. When faced with the lack of data, it is important to ask what can be learned about the system of interest from existing data.

If pertinent data do not exist for the system of interest, some things could be learned about its behavior by examining the behavior of similar systems. Medical researchers test the safety of new drugs on rats and monkeys because they are physiologically similar to humans. The concept of similarity can also be applied to chemical engineering. In the area of heat transfer, correlations for tubes are used for other geometries that are not drastically different from circular cross sections by replacing the diameter with an equivalent diameter. The condition "not drastically different" implies similarity.

Similarity is usually treated informally as engineering judgment. In order to apply the concept of similarity to analysis in a systematic way, a formal treatment of the concept is required. Kubic and Stein (1986) treated the similarity of fluid mixtures as a fuzzy binary relation and applied the concept to the problem of estimating uncertainty in physical property models. Their discussion, however, neglects the details of the similarity relation; and their application of similarity is limited. This paper expands on the work of Kubic and Stein by presenting a more rigorous and general development of similarity and by considering additional applications.

System Similarity

The concept of similarity can be described by fuzzy set theory. To begin the discussion, some *mathematical preliminaries* are required. Consider an ordinary set X called the universe. Membership in the ordinary subset A of X can be expressed by a membership function, μ_A, which is a mapping from X to $\{0, 1\}$ such that

$$\mu_A(x) = \begin{cases} 1 & \text{iff } x \in A \\ 0 & \text{iff } x \notin A \end{cases} \tag{1}$$

Zadeh (1965) introduced the concept of fuzzy subsets by allowing the membership function to have values on the interval $[0, 1]$. The closer $\mu_A(x)$ is to unity, the more x belongs to A. Hence, $\mu_A(x)$ defines a set with no sharp boundaries.

Ordinary set operations can be extended to fuzzy sets. Operations on fuzzy sets are defined by rules that give the membership function for each member of the resulting set. Some basic operations are given below.

$$\text{Union } (A \cup B) \quad \mu_{A \cup B}(x) = \max\,[\mu_A(x), \mu_B(x)] \tag{2}$$

$$\text{Intersection } (A \cap B) \quad \mu_{A \cap B}(x) = \min\,[\mu_A(x), \mu_B(x)] \tag{3}$$

Cross Product ($A_1 \times \cdots \times A_r$)

$$\mu_{A_1 \times \cdots \times A_r}(x_1, \cdots, x_r) = \min\,[\mu_{A_1}(x_1), \cdots, \mu_{A_r}(x_r)] \tag{4}$$

In ordinary set theory, a binary relation, xRy, defines a set of ordered pairs (x, y). R is the relation between x and y. A fuzzy binary relation defines a fuzzy set of ordered pairs. The membership function for each pair, $\mu_R(x, y)$, measures the strength

---

Correspondence concerning this paper should be addressed to F. P. Stein.

of the relation between x and y. To illustrate this concept, consider the expressions $x \gg y$ and $x \ll y$. These expressions are fuzzy relations because they represent an imprecise relationship between x and y.

System similarity can be defined loosely as a fuzzy binary relation describing physical systems that are approximately the same. System similarity is reflexive because a system is similar to itself. It should also be symmetric. Symmetry implies that if x is similar to y, then y is similar to x. The properties of reflexivity and symmetry place the following constraints on the membership function.

1. Reflexive $\mu_S(x, x) = 1 \quad \forall x \in \mathrm{dom}\, S$
2. Symmetric $\mu_S(x, y) = \mu_S(y, x) \quad \forall x, y \in \mathrm{dom}\, S$

A fuzzy relation that is reflexive and symmetric is a proximity relation (Dubois and Prade, 1980).

The behavior of a physical system is governed by the laws of chemistry and physics. These laws give rise to a set of physical attributes or parameters that uniquely characterize the system. Two systems behave the same if the physical attributes are the same because the physical phenomena are the same. The differences between two systems can be characterized by differences in the attributes. The less the differences, the greater the similarity.

Differences between systems can be described in a metric space (Groetsch, 1980). A metric space is defined by the pair (X, d) where X is a set and d is a nonnegative, real-valued function on $X \times X$ satisfying the following properties for any x, y, and z in X.

1. $d(x, y) = 0$ iff $x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, y) \leq d(x, z) + d(z, y)$

The metric, $d(x, y)$, is a generalization of distance. When applied to the problem of system similarity, X is the set of all pertinent systems; and the metric, d, measures differences between the systems.

A proximity relation and a metric space are related by the following proposition.

> *Proposition 1.* Let X be a set and $d(x, y)$ be a metric on $X \times X$. If $f(d)$ is a function such that
>
> 1. $f(0) = 1$
> 2. $0 \leq f(d) \leq 1$
>
> then $f[d(x, y)]$ defines a proximity relation on $X \times X$ with membership function $\mu_P(x, y) = f[d(x, y)]$.
>
> *Proof.* To be a membership function, $\mu_P(x, y)$ must be a mapping from $X \times X$ to $[0, 1]$. This requirement follows from condition 2. By definition, $d(x, x) = 0$; so by condition 1, $\mu_P(x, x) = 1$ proving $\mu_P(x, y)$ is reflexive. $d(x, y) = d(y, x)$, which implies that $\mu_P(x, y) = \mu_P(y, x)$, proving symmetry. $\mu_P(x, y)$ satisfies reflexivity and symmetry; therefore, it is a proximity relation.

As stated earlier, systems are similar if differences between them are small. This concept is expressed formally in the following definition of system similarity.

> *Definition 2.* A system similarity relation is a proximity relation with a membership function that can be expressed as a nonincreasing function of a metric. A nonincreasing function is a function such that $f(a) \leq f(b)$ if $a > b$.

It is often convenient to decompose a system into independent subsystems. Each subsystem can be characterized by a subset of the system attributes. A system similarity relation based on the appropriate subset of attributes can be used to compare subsystems. It should be noted that two systems are similar if their subsystems are similar. Independent subsystem similarity relations can therefore be combined to give the system similarity relation.

$$S = S_{sub,1} \times \cdots \times S_{sub,2} \tag{5}$$

Determination of a System Similarity Relation

To develop a similarity relation, one must identify the system attributes, construct a metric that reflects the relative importance of the attributes, and construct a membership function from this metric. This process involves a subjective element in the form of expert knowledge. In a sense, a similarity relation is not fundamentally different from engineering judgment. The formalism of similarity, however, offers several advantages. A similarity relation summarizes expert knowledge in a form that can be used easily by the nonexpert; it allows one to manipulate expert knowledge mathematically and to use the expert knowledge in engineering analysis; and it allows one to combine the knowledge of experts in different fields.

The system attributes required to develop a system similarity relation are obtained both from theoretical considerations and from a qualitative understanding of the system. Consider, for example, the single-phase flow of an incompressible liquid. The dimensionless Navier-Stokes equation indicates that the flow is completely characterized by Reynolds number, Froude number, and dimensionless geometry. Hence, these quantities form the basis for comparing liquid flow systems. In considering physical properties, molecular thermodynamics indicates that the attributes for comparing fluid mixtures are temperature, pressure, composition, and the chemical structure of each component. As with the fluid-flow example, attributes should be dimensionless. Because there is no completely rigorous theory of fluid mixtures, the dimensionless form of the attributes can only be approximate, which introduces a subjective element into the analysis.

The metric used in the similarity relation must satisfy the physical boundary conditions of the system. For large values of Reynolds numbers, fluid flow is independent of Reynolds number. A metric describing differences in Reynolds number must be consistent with the fact that differences in Reynolds number are not important for inviscous flow. In comparing the physical properties among a homologous series of molecules, the metric must reflect the fact that differences in molecular weight become less important as molecular weight increases. The difference between n-hexadecane and n-pentadecane, for example, are smaller than the differences between propane and ethane

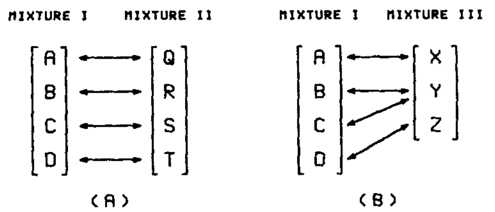

Figure 1. Molecular similarity of fluid mixtures.
(A) Equal number of components
(B) Unequal number of components
Arrows indicate similarity

although the absolute difference in molecular weight is the same.

The final step in constructing a similarity relation is to assign a value of the membership function to each value of the metric. This assignment depends on the importance one ascribes to the differences between systems and therefore is subjective.

The mechanics of determining a metric and a similarity relation are best studied using an illustrative example. In this study, fluid property similarity will be considered. The attributes of mixture similarity can be partitioned into four subsystems: temperature, pressure, composition, and molecules in the mixture. By Eq. 5, mixture similarity can be expressed as follows.

$$S_{mix} = S_{mol} \times S_{comp} \times S_T \times S_P \tag{6}$$

For simplicity, only molecular similarity will be considered.

Molecular similarity of a mixture is not easy to define. If molecule A is similar to molecule Q and molecule B is similar to molecule R, then the interactions between A and B should be similar to the interactions between Q and R. Based on this observation, it will be postulated that mixtures I and II are similar if for every component in mixture I there is a similar component in mixture II and for every component in mixture II there is a similar component in mixture I. This concept is illustrated in Figure 1a. Mixture I is similar to mixture III in spite of differing numbers of components because Y is similar to both B and C. This component-by-component comparison allows molecular similarity to be expressed as the cross product of component similarity.

$$S_{mol} = S_c \times \cdots \times S_c \tag{7}$$

Thus the problem of molecular similarity is simplified to a problem of component similarity.

Having developed a framework for the molecular similarity of fluid mixtures, the mechanics of developing a component similarity relation can be discussed. First, a metric is required to measure the differences between molecules. A structural comparison is an obvious choice, but it is very complex. An alternative is to choose measurable properties as attributes to characterize molecules. G. M. Wilson (private communication, 1985) uses such a scheme and compares fluids on the basis of normal boiling point and molecular weight. A property metric will be used in this study because of its simplicity.

To develop a property-based metric, the intrinsic dimensionality of fluid property space must be determined. Cramer (1980) applied factor analysis to fluid properties to determine the dimensionality of intermolecular interactions. In this study, factor analysis was used to study the correlation among the following properties: molecular weight, normal boiling point, critical temperature, critical pressure, critical volume, acentric factor, dipole moment, ideal gas heat capacity at 298 K, and heat of vaporization. The analysis was applied to all fluids in the data compilation of Reid et al. (1977). The results indicated that four independent factors characterize fluid-property space. Because the dimensionality of fluid-property space is four, four parameters or properties are required to characterize each molecule. The properties must span fluid-property space; they should be readily available for all molecules; and they should be relatively independent. Critical temperature, critical pressure, molecular weight, and reduced dipole moment satisfy the requirements and therefore are used as a basis of fluid-property space.

The following metric is proposed.

$$d(a, b) = \left[\left(\frac{T_{c,a} - T_{c,b}}{T_{c,a} + T_{c,b}}\right)^2 - a\left(\frac{P_{c,a} - P_{c,b}}{P_{c,a} + P_{c,b}}\right)^2 + b\left(\frac{M_a - M_b}{M_a + M_b}\right)^2 + c(\mu_{r,a} - \mu_{r,b})^2\right]^{1/2} \tag{8}$$

where $\mu_r = 10^5 \mu_p^2 P_c / T_c^2$.

This expression satisfies reflexivity, symmetry, and triangular inequality plus some intuitive properties of molecular similarity. The relative differences in the critical properties and molecular weight are more characteristic of molecular differences than absolute differences. Reduced dipole reflects the fact that although water and hexadecanol have about the same dipole moment, polarity does not contribute as much to the behavior of hexadecanol. Finally, it should be noted that using the absolute difference in reduced dipole avoids an indeterminate form for two fluids with no dipole.

The membership function for component similarity was assumed to have the following form.

$$\mu_S = \begin{cases} 1 & \text{if } d \le d_{min} \\ (d_{max} - d)/(d_{max} - d_{min}) & \text{if } d_{min} < d < d_{max} \\ 0 & \text{if } d \ge d_{max} \end{cases} \tag{9}$$

Expert knowledge was used to determine d_{min} and d_{max} and the parameters a, b, and c in Eq. 8. The expert knowledge was obtained from twelve industrial and academic thermodynamicists who rated the degree of similarity of 252 randomly selected pairs of molecules. To distill this information, a value of the membership function was assigned to each degree of similarity. The values of d_{min} and d_{max} were determined from pairs of fluids that all of the experts agreed upon as being definitely similar or completely different. The parameters in the metric were determined by a regression analysis, which minimized the squared error between the survey of experts and the computed membership function. The parameters for the metric and membership function are given in Table 1. While the component similarity relation presented here is reasonable, it should be viewed as a first attempt at determining a similarity relation.

Summarizing the example, there are three steps in determining a system similarity relation: postulating functions for the metric and similarity relation membership function, surveying experts to obtain the degree of similarity of sample systems, and fitting the results of the survey to the postulated functions.

Application of Similarity

The primary applications of a system similarity relation are evaluating data, correlations, and models. Two applications are

Table 1. Parameters for Property-Space Metric and Component Similarity Relation

a	1.62
b	0.44
c	8.3×10^{-6}
d_{min}	0.06
d_{max}	0.30

considered in this study: determining whether a correlation is applicable and evaluating the error in a model or correlation.

Correlation confidence

An engineering correlation is a mathematical representation of experimental data. From a strictly empirical point of view, there is no reason to expect a correlation to be applicable to a system that is not in the base set of experimental data; but one would expect reasonable predictions from a correlation if it were applied to systems that are similar to those contained in the base set. This observation suggests that a similarity relation can be used to measure the applicability of a correlation.

Let m be the system of interest and $\boldsymbol{D}$ be the set of data on which the correlation is based. If m is not in $\boldsymbol{D}$, one would expect the correlation to work if this system is similar to some system in $\boldsymbol{D}$. Systems similar to m are determined as follows.

$$\boldsymbol{M} = S[m] \tag{10}$$

The systems similar to m that are also in $\boldsymbol{D}$ are found by taking the intersection of $\boldsymbol{M}$ with $\boldsymbol{D}$. Because this intersection is a fuzzy set, all of its members are not equally similar to the system of interest. Conclusions about the applicability of the correlation to m should be based on the data set member that is most similar to m. A measure of correlation confidence is therefore defined as the maximum membership function of the intersection of $\boldsymbol{M}$ and $\boldsymbol{D}$, and it is computed as follows.

$$\mathcal{C}_m = \sup \min \left[\mu_s(x, m), \mu_D(x)\right] \tag{11}$$

For an untested correlation, the correlation data base, $\boldsymbol{D}$, contains only the data used to develop the correlation; and it represents the available knowledge about the correlation. As the correlation is tested, knowledge about the correlation increases, and $\boldsymbol{D}$ should be expanded to include all systems for which the correlation predictions are acceptable. Expanding the correlation data base is equivalent to saying that a correlation will work if it works for similar systems. It should be noted that if the correlation is applied to a system which is not similar to a member of $\boldsymbol{D}$, the predictions will not necessarily be poor. Lack of similarity may indicate that the correlation has never been tested on a particular class of systems, so accuracy cannot be guaranteed.

Knowledge of a correlation includes knowledge of systems for which the correlation fails as well as systems for which it works. If $\boldsymbol{F}$ is the set of systems for which the correlation fails, the intersection of $\boldsymbol{F}$ with $\boldsymbol{M}$ gives the fuzzy set of systems similar to the system of interest for which the correlation is known to fail. The existence of common members in $\boldsymbol{F}$ and $\boldsymbol{M}$ indicates that the correlation is not applicable to system m because the correlation does not work for a similar system. As with correlation confidence, a potential for failure is defined as the maximum membership function of the intersection of $\boldsymbol{M}$ and $\boldsymbol{F}$.

$$\mathcal{F}_m = \sup_x \min \left[\mu_s(x, m), \mu_F(x)\right] \tag{12}$$

In general, $\boldsymbol{D}$ and $\boldsymbol{F}$ are fuzzy sets. A correlation is tested by comparing predictions to data. If the errors in the predictions are acceptably small, the system is a member of $\boldsymbol{D}$; otherwise, it is a member of $\boldsymbol{F}$. The fuzziness of $\boldsymbol{D}$ and $\boldsymbol{F}$ arises from the fact that the size of an acceptable error is not well defined.

If model confidence is high and failure possibility is low, one expects predictions from the correlation to be accurate. If the opposite is true, predictions from the correlation are expected to be poor. When the possibility of success and failure are both large, there is a strong potential for both large and small errors. Low values for both correlation confidence and failure possibility indicate that the system of interest is not similar to any system in either $\boldsymbol{D}$ or $\boldsymbol{F}$. Hence, the correlation has never been tested on a similar system; so no statements can be made about its accuracy.

Engineering correlations usually have some theoretical basis. For example, heat transfer correlations are expressed as relations between dimensionless groups that can be determined from the energy and material balance equations. The theoretical element of a correlation enables the results to be extended to other systems for which the modeling assumptions are valid. Based on similarity arguments, a correlation for the second virial coefficient of methane as a function of temperature should not be applicable to xenon. However, if the critical pressure and critical temperature are used as reducing parameters to express the correlation in a dimensionless form, it would also be applicable to xenon because both methane and xenon can be described by two-parameter corresponding states. To summarize these observations, a correlation is applicable either if the model is valid for the system of interest, or if the system of interest is similar to the data on which the correlation is based.

The image of $\boldsymbol{D}$ through the similarity relation gives the set of systems similar to the systems in $\boldsymbol{D}$.

$$\boldsymbol{A} = S[\boldsymbol{D}] \tag{13}$$

Set $\boldsymbol{A}$ represents the limits of the empirical knowledge of the correlation. Let $\boldsymbol{C}$ be the set of systems for which the model assumptions are valid. The systems for which the correlation is valid are expressed mathematically as the union of $\boldsymbol{C}$ and $\boldsymbol{A}$.

$$\boldsymbol{V} = \boldsymbol{C} \cup \boldsymbol{A} \tag{14}$$

This relation is illustrated in Figure 2. The membership function for $\boldsymbol{V}$ measures the applicability of the correlation. The closer $\mu_V(x)$ is to unity, the greater the validity of the correlation. The membership function for $\boldsymbol{C}$ is determined using the methods of Kubic and Stein (1986). The membership function for $\boldsymbol{V}$ is computed as follows.

$$\mu_V(x) = \max \left[\mu_C(x), \mu_A(x)\right] \tag{15}$$

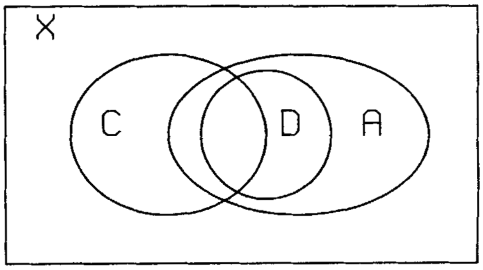

Figure 2. Regions of model applicability.

X = universe of systems
A = set of systems similar to correlation data base
C = set of systems for which model is valid
D = data base

Example

To illustrate the computation of correlation confidence, consider the use of the Peng-Robinson (1976) equation of state for predicting vapor pressures:

$$P = \frac{RT}{v - b} - \frac{a(T)}{v^2 + bv - b^2} \tag{16}$$

The parameter a, which was empirically determined from the vapor pressure data of 17 light hydrocarbons and inorganic gases, is a function of reduced temperature and acentric factor. It will be assumed for illustrative purposes that the only information available on this equation of state is the original paper. Hence, $\boldsymbol{D}$ contains only the 17 light hydrocarbons and gases. This example considers the application of the Peng-Robinson equation to six fluids that are not in the original data base. These fluids are *m*-xylene, methylamine, chlorotrifluoromethane, *n*-hexadecane, sulphur trioxide, and water.

To compute correlation confidence, a similarity relation is required. Only component similarity need be considered in this example. Because pure fluids are being studied, composition similarity is guaranteed, and molecular similarity reduces to component similarity. Temperature and pressure similarity are not important because a wide range of temperatures and pressures is being considered.

Table 2 gives the membership function from Eqs. 8 and 9 for similarity between the test fluids and the fluids used by Peng and Robinson to develop the correlation. By Eq. 11, the maximum membership in each column is the correlation confidence for the corresponding fluid. The correlation confidence for *m*-xylene and methylamine is unity, indicating that the correlation predictions should be good. The correlation confidence for chlorotrifluoromethane is 0.720, thus the Peng-Robinson equation can be used with a fair degree of confidence. Correlation confidence is less than 0.5 for the remaining three fluids. When these fluids are compared with the fluids in the Peng-Robinson data base, the differences are greater than the similarities. The lack of similarity does not necessarily mean that predictions will be poor, but that the correlation is being applied beyond the realm of experience, so accuracy cannot be guaranteed.

Table 2. Similarity of Example Fluids to Fluids in Peng-Robinson Equation Data Base

	Test Fluids					
Data Base	*m*-C_8H_{10}	CH_3NH_2	$CClF_3$	*n*-$C_{16}H_{34}$	SO_3	H_2O
N_2	0	0	0	0	0	0
H_2S	0	1.000	0	0	0.339	0
CO_2	0	0.562	0	0	0.453	0
CH_4	0	0	0	0	0	0
C_2H_6	0	0.209	0.110	0	0	0
C_3H_8	0.132	0	0.526	0	0	0
n-C_4H_{10}	0.621	0	0.681	0	0	0
i-C_4H_{10}	0.575	0	0.720	0	0	0
n-C_5H_{12}	0.966	0	0.632	0	0	0
i-C_5H_{12}	0.937	0	0.667	0	0	0
C_6H_6	0.785	0	0.304	0	0	0
cyclo-C_6H_{12}	1.000	0	0.481	0	0	0
n-C_6H_{14}	1.000	0	0.430	0	0	0
n-C_7H_{16}	1.000	0	0.233	0	0	0
n-C_8H_{18}	0.767	0	0	0	0	0
n-C_9H_{20}	0.598	0	0	0.219	0	0
n-$C_{10}H_{22}$	0.342	0	0	0.498	0	0
C_m	1.000	1.000	0.720	0.498	0.453	0
Avg. A lbs. Error (%) in V.P.	2.11	1.04	0.51	12.0	13.4	1.98

Average errors in vapor pressures are given in Table 2 for the six example fluids. Errors for *m*-xylene, methylamine, and chlorotrifluoromethane are small, which is expected because of the large value of correlation confidence. The large errors for *n*-hexadecane and sulphur trioxide are consistent with low correlation confidence. The vapor pressure predictions are surprisingly good for water. Although water is completely different from the fluids in the data base, the results are still consistent with the theory of correlation confidence. Low correlation confidence does not necessarily imply poor predictions, rather it indicates a potential for poor results because the correlation is being extrapolated beyond experience.

By testing the Peng-Robinson equation on the six fluids studied in this example, knowledge of the capabilities of this equation of state has increased. This knowledge can be used to augment $\boldsymbol{D}$ and to begin the construction of $\boldsymbol{F}$. Because the Peng-Robinson equation works for *m*-xylene, methylamine, chlorotrifluoromethane, and water, they should be added to set $\boldsymbol{D}$. Sulphur trioxide and *n*-hexadecane should be added to set $\boldsymbol{F}$.

Estimating correlation uncertainty

If correlation confidence is low and no data exist for the system, a similarity relation can be used to estimate the error in the correlation from existing data. The method for this procedure was developed by Kubic and Stein (1986). Their basic postulate is that when a model is applied to similar systems, the same error is obtained.

Set $\boldsymbol{Z}$ is the fuzzy set of systems similar to y, which is the system of interest.

$$\boldsymbol{Z} = S[y] \tag{17}$$

The intersection of $\boldsymbol{Z}$ with the set of existing data, $\boldsymbol{G}$, gives a set of similar systems for testing the correlation.

$$\boldsymbol{H} = \boldsymbol{Z} \cap \boldsymbol{G} \tag{18}$$

Let the elements of $\boldsymbol{H}$ be designated h. Associated with each of these elements is a set of errors obtained from comparison of predictions and data. The pairing of errors with systems can be represented by an ordinary relation, $\boldsymbol{hRe}$. The fuzzy set of errors, $\boldsymbol{E}$, is determined using the extension principle (Zadeh, 1975).

$$\mu_E(e) = \sup_{\substack{h \in H \\ hRe}} \mu_H(h) \tag{19}$$

Example

To illustrate the application of this method, consider the question of whether methanol and propene form a minimum boiling azeotrope at 25°C. This particular problem arises in a commercial ethylene plant. Methanol, which is used as a deicing agent, is often found in C_3-splitters. If methanol forms an azeotrope with propene, it will contaminate the propene product.

Table 3. Degree of Similarity of Some Binary Mixtures to Methanol/Propene

Components Similar to Propene	Components Similar to Methanol		
	CH_3OH $\mu = 1.000$	C_2H_5OH $\mu = 0.416$	1-C_3H_7OH $\mu = 0.00$
1,3-C_4H_6, $\mu = 0.746$	0.746*	0.416	0
i-C_4H_8, $\mu = 0.637$	0.637*	0.416	0
n-C_4H_{10}, $\mu = 0.496$	0.496*	0.416	0
trans-1,3-C_5H_8, $\mu = 0.261$	0.261*	0.261	0
n-C_5H_{12}, $\mu = 0.029$	0.029	0.029*	0
i-C_5H_{12}, $\mu = 0.052$	0.052	0.052*	0
C_6H_6, $\mu = 0.025$	0.025*	0.025*	0*

*Data in DECHEMA (Gmehling and Onken, 1977)

The criterion for the formation of a minimum boiling azeotrope can be expressed in terms of the infinite dilution activity coefficient of the less volatile component. An azeotrope will form if

$$\gamma_1^\infty \geq \phi_1^\infty P_2^s / \phi_1^s P_1^s \tag{20}$$

Because no data exist for the methanol/propene system at dilute methanol concentrations, activity coefficients must be estimated using a correlation. UNIFAC (Fredenslund et al., 1975) was used in this example.

A fuzzy set of binary mixtures similar to the methanol/propene system for which data exist was constructed. The set of similar mixtures was found by taking the cross product of the set of fluids similar to methanol and the set of fluids similar to propene. This operation is illustrated in Table 3. The membership function for the mixture is simply the minimum of the component membership functions.

The DECHEMA catalogs (Gmehling and Onken, 1977) were used as the data base for this example. Errors were expressed as the log of the ratio of the true value to the calculated value of the activity coefficient at infinite dilution.

$$e = \ln(\gamma_{data}^\infty / \gamma_{UNIFAC}^\infty) \tag{21}$$

The log ratio error is consistent with the constraint of positive activity coefficients. The membership function for the fuzzy set of errors is obtained from Eq. 19. This equation can be solved graphically by plotting the membership function of each similar mixture with its associated errors and drawing the smallest envelope that encloses all of the points. The membership function for this example is plotted in Figure 3.

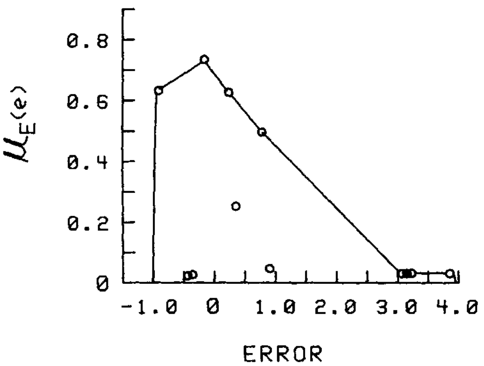

Figure 3. Fuzzy set of infinite dilution activity coefficient errors for methanol in propene as predicted by UNIFAC.

0 Error-membership function pairs as determined from data
– Membership function for the fuzzy set of errors

To determine whether or not an azeotrope will form, one must determine the possibility that the inequality in Eq. 20 is satisfied and the possibility that it is violated. The righthand side of Eq. 20 is 45.8 for the methanol/propene system at 25°C. UNIFAC predicts the infinite dilution activity coefficient of methanol to be 26.5, which indicates that no azeotrope will form.

Zadeh (1978) relates the potential for an event to occur to its membership function in a fuzzy set. The potential for forming an azeotrope can be determined by considering the true value of the infinite dilution activity coefficient to be a function of the error and applying the extension principle (Zadeh, 1975) to Eq. 20.

$$\Pi(\text{Azeotrope}) = \sup_{\gamma_m^\infty(e) \geq 45.8} \mu_E(e) \tag{22}$$

The solution to this equation is the maximum value of the membership function for which the error is greater than 0.55 (ln 45.8/26.5). Note that if the error is greater than 0.55, the true value of the infinite dilution activity coefficient is greater than 45.8, so an azeotrope will form. Thus, the potential for azeotrope formation is 0.56. Likewise, the potential for no azeotrope is as follows.

$$\Pi(\text{No azeotrope}) = \sup_{\gamma_m^\infty(e) < 45.8} \mu_E(e) \tag{23}$$

The potential for no azeotrope is 0.75.

Both the potential for forming an azeotrope and not forming an azeotrope are significant, which means that the correlation is not accurate enough to predict the formation of an azeotrope. The inaccuracy of the correlation indicates that data are required for this system. The authors' experience with ethylene plants indicates that methanol and propene do form an azeotrope. Application of the similarity relation to this example indicates the potential for a serious problem with the azeotrope. UNIFAC does not suggest any problem; it clearly predicts no azeotrope.

Conclusions

A detailed development of system similarity has been presented in this paper. Methods for determining a system similarity relation were illustrated. A method for measuring correlation confidence and the possibility of correlation failure based on a similarity relation was developed. Finally, a method for estimating the error in a correlation was discussed.

The power of a system similarity relation is that it is a method of representing expert knowledge that can be easily manipulated and automated. If the proper similarity relations are available, measures of correlation confidence can be included as part of a simulation package to inform the user when predictions may not be reliable.

Notation

a = Peng-Robinson attractive parameter
A = set of systems similar to the elements of D
b = Peng-Robinson hard-sphere parameter

$\mathbf{C}$ = set of conditions for which a model is valid
$\mathcal{C}_m$ = correlation confidence for system m
$d(\cdot)$ = metric
d = element of set $\mathbf{D}$
$\mathbf{D}$ = correlation data base
e = error
$\mathbf{E}$ = fuzzy set of correlation errors
f = element of set $\mathbf{F}$
$\mathbf{F}$ = set of systems for which a correlation fails
$\mathcal{F}_m$ = potential of correlation failure for system m
$\mathbf{H}$ = set of similar systems for which data exist
M = molecular weight
$\mathbf{M}$ = fuzzy set of systems similar to system m
P = pressure
P^s = vapor pressure
R = gas constant
$\mathbf{R}$ = error relation
$\mathbf{S}$ = system similarity relation
T = temperature
v = molar volume
$\mathbf{V}$ = set of systems for which a correlation is valid
x = elements of set $\mathbf{X}$
$\mathbf{X}$ = universe
$\mathbf{Z}$ = set of similar systems

Greek letters

γ = activity coefficient
ϕ = fugacity coefficient
$\mu(\cdot)$ = membership function
μ_p = dipole moment
μ_r = reduced dipole

Superscripts

s = saturated property
∞ = infinite dilution

Subscripts

c = critical property
$comp$ = composition
$data$ = data
m = methanol
mix = mixture
mol = molecular
P = pressure
T = temperature
$UNIFAC$ = UNIFAC

Literature Cited

Cramer, R. D., "BC(DEF) Parameters. 1: The Intrinsic Dimensionality of Intermolecular Interactions in the Liquid State," *J. Am. Chem. Soc.*, **102,** 1837 (1980).

Dubois, D., and H. Prade, *Fuzzy Sets and Systems: Theory and Applications,* Academic Press, New York (1980).

Fredenslund, A., A. L. Jones, and J. M. Prausnitz, "Group-Contribution Estimation of Activity Coefficients in Nonideal Liquid Mixtures," *AIChE J.*, **21,** 1086 (1975).

Groetsch, C. W., *Elements of Applicable Functional Analysis,* Dekker, New York (1980).

Gmehling, J., and U. Onken, *Vapor-Liquid Equilibrium Data Collection,* **1,** DECHEMA, Main, BRD (1977).

Knapp, H., "Physical Properties in Process Design—Past, Present, and Future," *Fluid Ph. Equil.*, **30,** 1 (1986).

Kubic, W. L., and F. P. Stein, "The Application of Fuzzy Set Theory to Uncertainty in Physical Property Models," *Fluid Ph. Equil.*, **30,** 111 (1986).

Peng, D., and D. B. Robinson, "A New Two-Constant Equation of State," *Ind. Eng. Chem. Fundam.*, **15,** 59 (1976).

Reid, R. C., J. M. Prausnitz, and T. K. Sherwood, *The Properties of Gases and Liquids,* McGraw-Hill, New York (1977).

Wilson, G. M., Private Communication (1985).

Zadeh, L., "Fuzzy Sets," *Information and Control,* **8,** 333 (1965).

———, "The Concept of a Linguistic Variable and Its Application to Approximate Reasoning, 1," *Inf. Sci.*, **8,** 199 (1975).

———, "Fuzzy Sets as a Basis for a Theory of Possibility," *Fuzzy Sets and Systems,* **1,** 3 (1978).

Manuscript received Feb. 13, 1987, and revision received July 13, 1987.