ChE classroom

THE SYNTHETIC-DATA METHOD

WALLACE B. WHITING, HUI-MIN HOU,[1] SHAO-HWA WANG[2]
West Virginia University
Morgantown, WV 26506

Thermodynamics; statistics; process design. Few courses are viewed with as much trepidation by chemical engineering students (and faculty) as these are. Yet we generally admit that more understanding of these topics (even if not more courses) is essential to the success of our students.

The synthetic-data method provides a framework for integrating these fields and thereby making them more "real" for chemical engineering students. The fundamental principle of the method is optimization: making the most and best use of limited experimental data. As such, it involves error analysis, statistics, thermodynamics, and process design.

SYNTHETIC-DATA METHOD

We are often faced with the problem of too few experimental data and too simplistic models in chemical engineering; a classic example is fluid-phase equilibria. Our measurements of fluid systems (temperature, pressure, composition, etc.) are quite good, but our models are simplistic: cubic equations of state, activity-coefficient models, etc. A further complicating aspect of the problem is that we have too few of these experimental measurements. It is not feasible, even for common systems, to have complete physical property data at all temperatures, pressures, and compositions of interest. There are no data at all for many systems. Over the years, a set of procedures has been developed to solve this problem—the synthetic-data method.

The general idea is to generate artificial (synthetic) data for the system of interest from group-contribution or other methods. One then regresses these synthetic data to determine the parameters in the thermodynamic models that one wishes to employ. Group-contribution techniques are not new, and we routinely expose undergraduates to, for example, the Lydersen technique[1] for estimating critical temperatures and pressures from the molecular structure of the compound. Of course, it is not the critical constants that are important—they are merely synthetic data. But, from these, we derive the set of parameters that we need for our equation of state.

***Wallace B. Whiting** is Associate Professor of Chemical Engineering at West Virginia University, where he has taught for the past decade. He is active in ASEE and AIChE, and his research and teaching interests range from thermodynamics to process safety and process design.*
***Hui-Min Hou** received her degrees from West Virginia University (MSChE) and Taiwan National University (BSChE). She has had broad experience in her present position at PLG, Inc., as well as previously at Halliburton NUS. Her specialty is chemical process risk assessment.*
***Shao-Hwa Wang** received his degrees from West Virginia University (PhD, MSChE) and Taiwan National University (BSChE). Previously at M.W. Kellogg, and now in the Process Economics Program at SRI International, Dr. Wang's work has ranged from thermodynamics to process*

An increasingly important technique in industry is the use of a group-contribution activity-coefficient technique (UNIFAC) to generate synthetic vapor-liquid equilibrium data, which are then regressed to determine equation-of-state parameters.[2] Many applications and variations of this technique have been reported in the literature.[3,4] Recently, a related approach was presented[5] in which very limited infinite-dilution activity coefficient data plus the Wilson equation are used rather than the group-contribution idea to create synthetic data sets for regression of equation-of-state parameters.

The steps in the synthetic-data method are shown schematically in Figure 1 and are described below.

- Determine the best available primitive model and the data available. When data are sparse (the usual case), a group-contribution technique is chosen. In our application we use the UNIFAC model for liquid-state activity coefficients.
- Generate synthetic data from the primitive model chosen. These data should be as close as possible to the range of conditions of interest in the problem to be solved, but they must be within the range of validity of the primitive model. Typi-

[1] PLG, Inc., 4590 MacArthur Blvd., Suite 400, Newport Beach, CA 92661-1017
[2] SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025

cally, group-contribution techniques are much more limited in application range than are the models that are needed to solve the problem. For example, the UNIFAC model is good only for low pressures and near-ambient temperatures.

- The parameters in the final model to be used are regressed from the synthetic data generated. In the regression, these data are weighted according to the needs of the problem. In our examples we use the Mathias version of the Soave-Redlich-Kwong equation of state.[6]

The synthetic-data method is powerful and adaptive. It is, in effect, a "bootstrap" procedure. From only the chemical structure of the substances in the mixture, data are created for one set of conditions. The parameters for the more general model are regressed from these synthetic data, and predictions of phase equilibria over a broad range of conditions are then made. The engineer chooses which synthetic data to use and how to weight them in the regression of the final model parameters. Thus, the higher levels of engineering judgment (analysis, synthesis, and evaluation) must be used by the engineer or the engineering student.

The importance of these synthetic-data methods in teaching is that they create a framework for the integration of thermodynamic models, experimental data, statistics, and process design.

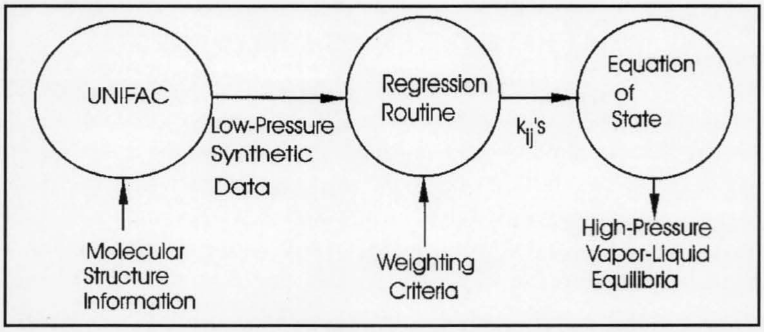

Figure 1. *Schematic example of the synthetic-data method.*

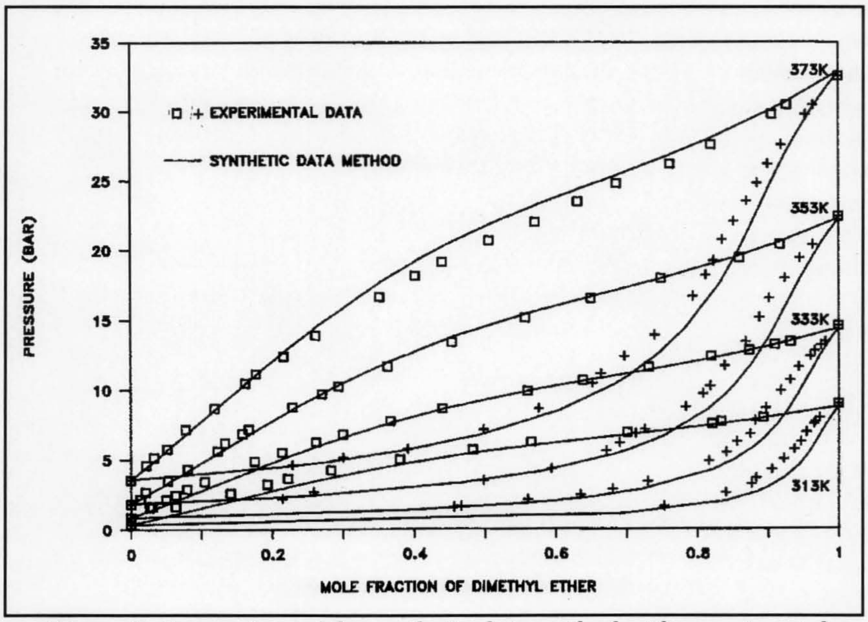

Figure 2. *Comparison of the synthetic-data method with experimental data for dimethyl-ether/methanol*

THERMODYNAMICS

We teach thermodynamics because we want students to understand its great unifying concepts: energy, mass, entropy, phase equilibrium, reaction equilibrium. But the test of that understanding in their profession is if they can use thermodynamic models for simulation of processes, whether or not the context of the assignment is plant operations, research, design, or sales. It is difficult to put these models into perspective in a short four-year curriculum. But it should be getting easier.

With user-friendly computer programs now available, our students can try different models, compare them to data, and experience the reality that these models, as elegant and complex as they may seem to be, are only crude approximations of reality and should be treated as such. The synthetic-data method is a good vehicle for this instruction.

The students are given a typical problem: they are asked to calculate the vapor-liquid equilibrium for a binary system of dimethyl-ether/methanol. (Any system may be chosen, but the results for this system are given in Figure 2.) To accomplish this, the students must choose a thermodynamic model, and they must know the parameters in that model. The choice of the model and the calculation of the compositions for which the fugacities are equivalent in the two phases are important, and nontrivial, assignments. The instructor provides the data. Of course, the data could be found easily in the literature (especially for this system), but we suggest that synthetic data be generated from, for example, the UNIFAC model and presented to the students. Depending on the students' backgrounds, we suggest that the ensuing parts of the problem be made more interesting by "errorizing" the data with a simple Gaussian distribution of "experimental" error.

The students submit their solutions, which should include the parameters that they have regressed, the vapor-liquid equilibria that they have calculated, and some measure of the deviations of the calculated results from the experimental data that were provided. During the discussion of their results, which should be a "reflection in action"[7] about what they have done, some or all of the following concepts can be brought in—concepts that would normally seem esoteric to the students but which are now of vital importance:

Experimental Error. • The instructor has introduced this artificially, but the students will be able to estimate (to varying degrees) what the experimental error was. The discussion can easily range from random to systematic errors, to the replication of experiments, to techniques for evaluating which model is best, to consequences of inaccurate model predictions, to sources of experimental data.

Statistics • Many people (including ABET and industrial advisory committees) decry the lack of statistical understanding of our students. But clearly the solution is not to ship the students off to mathematics or statistics departments for the types of courses that have created fear and anxiety about statistics in generations of students. Why not *use* statistics in existing courses? Chemical engineering students have a compelling need for statistics in, for example, thermodynamics. One can discuss experimental error, quality of physical-property models, statistical significance of differences between them, confidence regions of the parameters, maybe even thermodynamics consistency in the context of statistics. If we want to be sure that students will have the motivation for this discussion, we can give them different sets of the binary data for the problem and have them compare their results with one another.

Choice of Thermodynamic Model • The very different results that students get from their chosen models naturally leads to this important discussion.

Synthetic-data method • At some stage in the discussion described above, the instructor can explain how the data were generated for the problem, and the discussion will quickly turn to an examination of the synthetic-data method: how it can be (and is) used; when it is an appropriate choice; what its limitations are. Asking students to come up with other examples of the synthetic-data method can lead to even more unifying discussion.

PROCESS DESIGN

The ubiquitous use of process simulation programs in chemical engineering design courses presents exciting opportunities for students to acquire experience. Again, we suggest the synthetic-data method as a unifying concept for acquiring this experience.

When students are designing a process, a major stumbling block is typically the thermodynamic model. Encouraging students to use the default model is dangerous and unnecessary.[8] Instead, we encourage students to choose the "best" model and give them synthetic data as described above. In this way, they use the regression skills they learned in previous courses as well as the thermodynamic concepts that they have mastered.

Each of the design groups chooses a different model, either on its own or through instructor encouragement. An active class discussion ensues in which the different designs of their process units are compared. The direction of this discussion follows the example given above for the thermodynamics class, but here the focus is not just on the disparity between the data and the vapor-liquid equilibria, but also on the apparent discrepancy between any of the designs and the actual operation of a real plant.

As was the case in the thermodynamics example, the final discussion here involves the students finding examples of the synthetic-data method—but this time they try to find examples in the various thermodynamic property options of the simulator.

WHAT HAVE THE STUDENTS LEARNED?

In the thermodynamics example, the students may not have learned what entropy is, and in the process design example they may not have learned anything about eigenvalues. But they certainly *have* learned about how to choose thermodynamic models, how important thermodynamics really is, and how much faith to have in the results. They have developed an expertise that they are likely to remember and to use when the need arises. Perhaps (we think definitely) they will have learned some statistics, again, in a way that they will remember and use.

CONCLUSION

The synthetic-data method provides a framework for unifying thermodynamics, process design, and statistics in such a way that students gain valuable experience in using the concepts they are learning.

ACKNOWLEDGEMENT

We appreciate the partial financial support of the U.S. Department of Energy through the Consortium for Fossil Fuel Liquefaction Science.

REFERENCES

1. Reid, R.C., J.M. Prausnitz, and B.E. Poling, *The Properties of Gases and Liquids,* 4th ed., McGraw-Hill Book Company, New York, NY (1987)
2. Wang, S.-H., and W.B. Whiting, "Group-Contribution Binary-Interaction Parameters for Equations of State," presentation at 1987 AIChE Spring National Meeting, paper 27f, Houston, TX, March (1987)
3. Schwartzentruber, J., and H. Renon, "Extension of UNIFAC to High-Pressures and Temperatures by the Use of a Cubic

Equation of State," *Ind. Eng. Chem. Res.*, **28**, 1049 (1989)
4. Santacesaria, E., R. Tesser, and M. Di Serio, "Simple and Predictive Approach for Calculating the High Pressure and Temperature Vapour-Liquid Equilibria of Binary Mixtures by Applying a UNIFAC Equation of State Method," *Fluid Phase Equil.*, **63**, 329 (1991)
5. Twu, C.H., D. Bluck, J.R. Cunningham, and J.E. Coon, "A Cubic Equation of State: Relation Between Binary Interaction Parameters and Infinite Dilution Activity Coefficients," *Fluid Phase Equil.*, **72**, 25 (1992)
6. Mathias, P.M., "A Versatile Phase Equation of State," *Ind. Eng. Chem. Proc. Des. Dev.*, **22**, 385 (1983)
7. Schön, D.A., *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions*, Jossey-Bass, San Francisco, CA (1987)
8. de Nevers, N., and J.D. Seader, "Helping Students Develop a Critical Attitude Towards Chemical Process Calculations," *Chem. Eng. Ed.*, **26**, 88 (1992) ❐

ChE book review

INTRODUCTION TO PHYSICAL POLYMER SCIENCE, 2nd Edition

by L.H. Sperling

John Wiley & Sons Inc., New York, NY; 594 pages, $64.95 (1992)

Reviewed by
Eric A. Grulke
Michigan State University

Polymer physical science (the combination of polymer physics and polymer physical chemistry) forms the basis for interpreting and solving a wide variety of polymer processing and polymer performance problems. The first edition of Sperling's *Introduction to Physical Polymer Science* provided a good introduction to the field for chemical engineers and material scientists alike. The second edition has been expanded in several important areas: the amorphous and crystalline solid states, liquid crystalline systems, and mechanical behavior. It is a valuable reference for industrial practitioners as well as a good introductory textbook.

The book begins with a short overview of polymers, followed by descriptions of chain structures and configurations, and molecular weight distributions. The middle chapters provide descriptions of concentrated solutions and polymer blends, the amorphous state, the crystalline state, liquid crystalline polymers, and thermalmechanical transitions. The final chapters cover mechanical and flow properties, including the elasticity of crosslinked polymers, polymer rheology and viscoelasticity, mechanical behavior, and some selected topics.

The introductory material in Chapter 1 provides the reader with an adequate background and vocabulary to read the rest of the text. Chapter 2 deals with chain structure and emphasizes stereochemistry, isomerism, copolymer types and morphologies, and photophysics. Descriptions of chain structure analytical methods provide an introduction to polymer characterization techniques.

Polymer molecular weight determinations are covered in Chapter 3. Polymer solution thermodynamics forms the basis for these measurements and is covered early in the chapter, an improvement from the first edition. Colligative, light scattering, solution viscosity, and gel permeation chromatography techniques are presented. The second edition includes worked example problems starting in Chapter 3—an important improvement for classroom use and self-study alike.

Phase separation behavior (Chapter 4) has received much better coverage in the second edition. There are additional phase diagrams, an expanded discussion of polymer-polymer miscibility, and a good summary of the kinetics of phase separation. The section on diffusion and permeability in polymers should be helpful to those interested in packaging applications.

The material on bulk states (amorphous and crystalline) has been expanded into separate chapters (Chapters 5 and 6) and a new chapter has been added on liquid crystals (Chapter 7). These changes have made this edition of *Introduction to Physical Polymer Science* one of the best single references for the physical science description of solid and solid-like polymer systems.

The discussion of amorphous polymers includes short-range interactions and long-range order, the conformation of the polymer chain and macromolecular dynamics. Two models for linear polymer motion are presented: a bead-and-spring model (Rouse-Bueche theory) and the reptation model (de Gennes). In addition, the motion of nonlinear chains is described.

Chapter 6 on the crystalline state includes analytical methods for determining crystal structure, unit cells, chain structures, crystallization from the melt, crystallization kinetics, and the thermodynamics of fusion. There are also good sections on the re-entry of chain segments in lamellae, the effect of chemical structure on the melting temperature, and fiber formation and structure.

Chapter 7 on the liquid crystalline state is new to this edition. There are sections on mesophase types and morphologies, fiber formation, comparison of major polymer types, and the requirements for liquid crystal formation.

The material on thermal-mechanical transitions (Chapter 8) and rubber elasticity (Chapter 9) is about the same as in the first edition. The five regions of viscoelastic behavior are explained well, and there is a good section on theories of the glass transition. There are three laboratory/lecture demonstrations that help illustrate concepts of rubber elasticity.

Continued on page 152.