

WORKFLOW AND REGRESSION METHODS FOR DETERMINING SOLID-LIQUID PHASE DIAGRAMS

Ketan D. Samant and Lionel O'Young
ClearWaterBay Technology, Inc.
20311 Valley Blvd, Suite C, Walnut, CA 91789

Martin Kwok and Ka M. Ng
Hong Kong University of Science & Technology
Clear Water Bay, Kowloon, Hong Kong

Abstract

Systematic procedures for synthesis and design of crystallization-based separations are all founded on understanding the solid-liquid phase behavior. Experimental and computational tools that give both quantitative results and qualitative insight into solid-liquid phase behavior are therefore very valuable. This paper provides a glimpse into our effort of combining the experimental and computational tools into an organized workflow for determination of solid-liquid phase diagrams. It also describes in detail one component of the workflow, regression methods.

Keywords

Crystallization, Solid-liquid phase diagrams, Regression, Workflow, Process synthesis, Conceptual design.

Introduction

Crystallization has traditionally served as a purification process in the chemical industry. However, with the recent shift of emphasis to high value-added products, interest in development of separation processes based on crystallization technology has increased dramatically. This is especially the case in the pharmaceutical and specialty-chemicals sectors.

Wibowo and Ng (2002) have identified and described in detail the key steps involved in the development of a crystallization-based separation process. The development typically begins in the laboratory. Various experiments are performed with the aim of understanding the properties of a new product and gaining some insight into the crystallization process. These experiments include DSC, TGA, PXRD, Raman, IR, etc. and the experimental data include solubility in various solvents, melting points, and information about solvates, hydrates, polymorphic forms, etc. The experimental data are then used for the iterative

process of designing and evaluating a reliable and optimal process with minimum time, effort, and money. The task of creating an accurate understanding of the underlying solid-liquid phase behavior is crucial to the success of this process. This is a very challenging task that is made even more difficult by the lack of a systematic workflow and the lack of reliable regression methods that can fit the experimental data to valid thermodynamic models.

We have been working on developing experimental and computational tools for determining solid-liquid phase diagrams for simple molecular systems, systems with polymorphs, systems with solvates and solid-complexes, systems with liquid phase reactions, and ionic systems. In this paper, we will begin with a brief overview of the workflow that we have developed during the course of this work. We will then describe our experience with methods for regression of solid-liquid solubility data. More specifically, with the help of an example, we will discuss

the limitation of traditional least-squares based regression methods when the available solubility data do not span the entire composition space. We will then present a modification to the traditional approach and show how it overcomes this limitation by introducing penalties derived from the solid-liquid stability criterion.

Workflow

The term workflow collectively represents a set of activities, resources required to perform these activities, and the flow of information between the activities and resources, all to achieve a clearly defined set of objectives. Fig. 1 shows the schematic of our workflow for determination of solid-liquid phase diagrams. The objective here is to create an accurate representation of the solid-liquid phase behavior of the system under consideration.

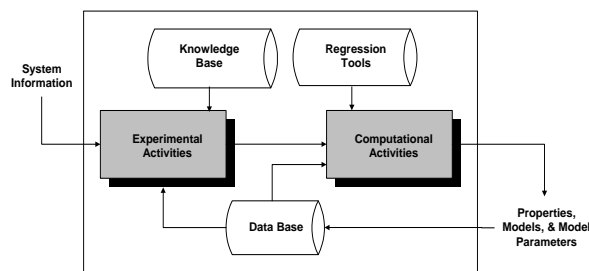


Figure 1: Workflow for Determination of Solid-Liquid Phase Diagrams

As shown in this figure, the workflow comprises two main activities. The experimental activities include the actual laboratory experiments. The computational activities include the tasks of digesting the laboratory data and expressing them in a form suitable for process synthesis and design. To effectively carry out these activities, three main resources are needed. These resources and their important components are listed in Table 1.

Table 1: Resources and Their Components

Resource	Components
Knowledge Base	Equipment and resources
	Experimental methods archive
	Experimental method selection tool
Regression Tools	Property models and routes
	Regression methods
Database	Experimental data tables
	Model parameter tables

The knowledge base helps in selecting experiments based on system type and variables to be measured and

also provides information about equipment and experimental procedures. The database provides access to past experimental data as well as to previously obtained model parameters. The regression tools provide valid thermodynamic models and regression methods that help express the experimental data into a meaningful form as needed.

Details on the activities, resources, and flow of information in this workflow can be found in Kwok (2003). In the rest of this contribution we will discuss some details related to regression methods.

Regression Methods

In this section we will focus our attention on some details of the regression methods. We will consider simple molecular systems to illustrate a particular limitation of traditional methods and present a solution to overcome this limitation. This limitation and its solution are applicable even for systems with polymorphs, adducts, and liquid phase reactions.

SLE Equations and Stability Criterion

The most commonly used form of solid-liquid equilibrium equation is as follows:

$$\frac{f_k^S}{f_k^L} = \gamma_k x_k \quad (1)$$

In this equation, the left-hand side is the ratio of fugacities of pure component k in solid and liquid phase at a specified temperature and pressure. The right-hand side represents the activity of component k in the liquid phase at the same conditions. This equation is also called the saturation variety equation for component k as it represents all liquid phase compositions at which this component is saturated. Note that the corresponding solid phase in equilibrium with this liquid is pure k . The solid-liquid phase diagram for a system is the intersection of the saturation varieties for all components in the system.

As derived by Michelsen (1982), stability of a mixture of composition \mathbf{x} , requires that

$$\sum_k y_k [\mu_k(\mathbf{y}) - \mu_k(\mathbf{x})] \geq 0 \quad (2)$$

for all trial compositions \mathbf{y} . Based on this equation, the condition for a liquid of composition \mathbf{x} to be stable with respect to formation of solid k can be derived as:

$$\frac{f_k^S}{f_k^L} \geq \gamma_k x_k \quad (3)$$

If this criterion is satisfied, the liquid is stable with respect to formation of a solid phase comprising of pure component k .

Traditional Approach

As Eqn. (1) indicates, the only unknown mixture-dependent properties in the SLE equation are the liquid phase activity coefficients. The objective of regression is therefore to obtain the interaction parameters that are typically used in activity coefficient models. The most commonly used activity coefficient models account only for binary interactions. Therefore, solid-liquid equilibrium experiments typically involve two components.

The binary SLE experiments are performed at a fixed pressure. At this pressure, the experimental data points are obtained by measuring the solubility (or the saturation composition) of one of the components at different temperatures. This collection of experimental data is used to obtain the binary interaction parameters by solving the following optimization problem:

$$\min F = \sum_{i \in \text{DataSet}} w_i (\gamma_{\text{solute}}^{\text{calc}} - \gamma_{\text{solute}}^{\text{exp}})_i^2 \quad (4)$$

In this equation, the suffix solute indicates the component whose solubility is measured at the data point being considered. The other component in the binary system is the solvent at this data point. The experimental value of the activity coefficient is obtained from Eqn. (1). With this substitution, the above regression optimization takes the following form:

$$\min F = \sum_{i \in \text{DataSet}} w_i \left(\gamma_{\text{solute}}^{\text{calc}} - \frac{f_{\text{solute}}^S}{x_{\text{solute}} f_{\text{solute}}^L} \right)_i^2 \quad (5)$$

Limitation of the Traditional Approach

A closer look at Eqn. (1) reveals that it only contains information about the saturation variety of component k . It does not provide any information about the saturation level of all other components in the system. However, this knowledge is necessary for accurate representation of the phase behavior because, as we have noted earlier, the phase behavior is actually the intersection of these saturation varieties.

Since Eqn. (1) is used in the regression problem formulation of Eqn. (5), this formulation also makes no attempt to take into account the saturation levels of the solvent component at each data point. It simply tries to faithfully reproduce the saturation variety of the solute component at each data point. Quite often this creates a severe problem with the resulting solid-liquid phase diagram although the regression optimization itself seems to give very accurate results.

Let us elaborate with an example of a simple binary system of components A and B. The experimental data for this system is summarized in Table 2. Results of the regression calculations performed according to the formulation of Eqn. (5) are shown in Fig. 2. The percent errors at each data point are listed in Table 3.

Table 2: Experimental Data for Example System

Data Point	Temperature (C)	Solubility of A (mole fraction)
1	27.0	0.896
2	24.5	0.795
3	22.8	0.690
4	21.0	0.583
5	20.0	0.529
6	19.7	0.497

The formulation of Eqn. (5) does an excellent job of fitting the saturation curve of component A, as can be seen from Table 3. However, the resulting phase diagram as shown in Figure 2 does not look as good. In this diagram, the curve on the RHS is the saturation curve for solute A. The curve on the LHS is the saturation curve for solvent B. The two curves intersect at the eutectic point marked as E1. Below the eutectic, the system exists only as a mixture of solid A and B. Clearly then, the last two data points (circled) appear incorrectly in the solids-only region of the phase diagram.

This result can be attributed to the fact that the formulation of Eqn. (5) simply tried to faithfully reproduce the saturation curve of A. It paid no attention to the saturation curve of B. The location of the eutectic point and the solids-only region depends on both these saturation curves. Therefore, they are incorrectly placed on the final phase diagram.

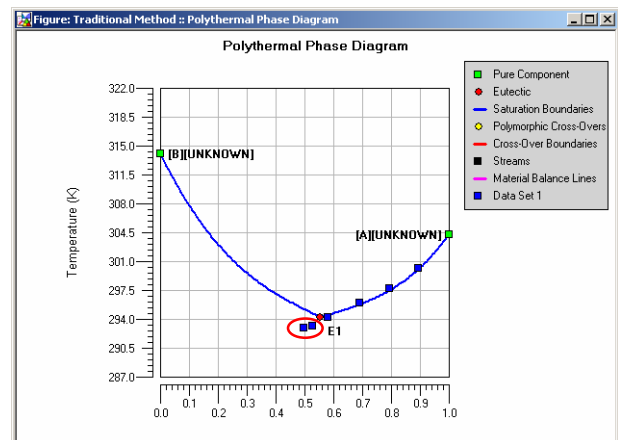


Figure 2: Phase Diagram with Traditional Formulation

Overcoming the Limitation

The best way to overcome this limitation is to obtain the solubility data spanning the entire composition space. However, this is easier said than done and in some situations may even be considered a wasteful allocation of resources.

The other solution is to modify the traditional regression formulation. We know that the problem is caused by the lack of knowledge about the saturation level of the solvent component. If we can mathematically incorporate this knowledge in the formulation, we will have solved the problem.

The stability criterion derived in Eqn. (3) provides us with this mathematical representation. At each data point, the liquid composition must be stable with respect to formation of solid phase comprising the solvent component. Therefore, at each data point, the solvent component must satisfy the stability criterion. Therefore, by adding a penalty term based on this criterion should force the regression formulation to give more accurate phase diagrams. The new formulation looks as follows:

$$\min F = \sum_{i \in \text{DataSet}} w_i \left(\gamma_{\text{solute}}^{\text{calc}} - \frac{f_{\text{solute}}^S}{x_{\text{solute}} f_{\text{solute}}^L} \right)_i^2 + \sum_{i \in \text{DataSet}} g \left(\frac{f_{\text{solvent}}^S}{f_{\text{solvent}}^L}, \gamma_{\text{solvent}}^{\text{calc}} x_{\text{solvent}} \right)_i \quad (6)$$

Let us now go back to the example of a binary system considered earlier. Fig. 3 shows the results obtained using the above formulation. The results for each individual data point are shown in Table 3. A quick comparison reveals that the addition of the stability-criterion based term does not affect the fit for the saturation curve of component A. It however does a better job of accurately predicting the saturation curve for component B and enhances the accuracy of the phase diagram considerably.

Table 3: Comparison of Regression Methods

Data Point	% Error	
	Traditional Method	New Method
1	0.053	0.051
2	0.109	0.104
3	0.331	0.325
4	-0.169	-0.173
5	-0.410	-0.410
6	0.330	0.333

Conclusions

In the past few years interest in development of separation processes based on crystallization technology has increased dramatically. This interest has generated a lot systematic procedures and tools for tackling synthesis and design issues for crystallization-based separations. These tools are all founded on understanding the solid-liquid phase behavior.

We believe that experimental and computational tools that give both quantitative results and qualitative insight into solid-liquid phase behavior are therefore very valuable and these tools need to be developed and combined into an organized and efficient workflow. In this short paper, we have presented a part of our work in this ongoing effort.

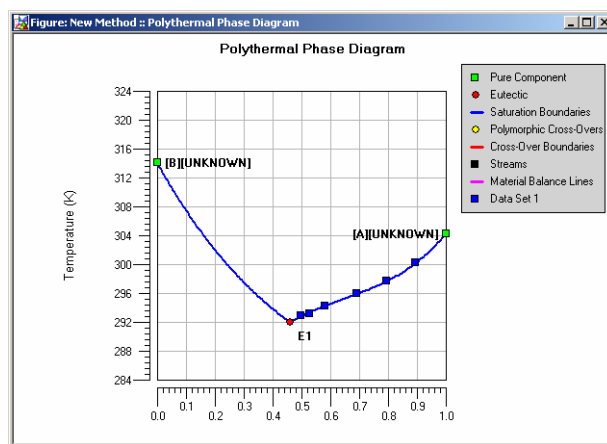


Figure 3: Phase Diagram with New Formulation

Nomenclature

- f = fugacity
- x, y, x, y = compositions, mole fraction
- w = weight
- γ = activity coefficient
- μ = chemical potential

References

- Kwok, K. S. (2003). *Experimental determination of solid-liquid equilibrium (SLE) phase diagrams of different organic-based chemical systems*. Unpublished MS dissertation, Hong Kong University of Science and Technology, Hong Kong.
- Michelsen, M. L. (1982). The isothermal flash problem. Part 1. Stability. *Fluid Phase Equilibria*, **9**, 1-19.

Wibowo, C. and K. M. Ng (2002). Workflow for process synthesis and development: crystallization and solids processing. *Ind. Eng. Chem. Res.*, **41**, 3839-3848.