Korean J. Chem. Eng., **26**(2), 349-353 (2009)

Prediction of the surface tension of binary systems based on the partial least squares method

Sung Young Kim, Sung Soo Kim, and Bomsock Lee†

Department of Chemical Engineering, Industrial Liaison Research Institute, KyungHee University, Yongin 446-701, Korea
(*Received 10 June 2008 • accepted 29 August 2008*)

Abstract−To predict the surface tension of binary liquid systems, an empirical model is proposed using the partial least squares (PLS) based on the multivariate statistical analysis method. Required parameters for the PLS method to predict the surface tension of binary systems are composed of the thermophysical properties of only pure substances such as critical temperature, critical pressure, critical volume, molar volume, viscosity and vapor pressure for input data block (X) and the reported experimental surface tension data for output data block (Y). The data set for the experimental surface tension of binary liquid systems is divided into the training set for regression and the test set for predicting. An average relative error (%) results of regression and prediction indicate that the PLS method can be a useful tool for predicting the surface tension of liquid binary systems.

Key words: Surface Tension, Partial Least Squares, PLS, Multivariate Statistical Analysis, Binary System

INTRODUCTION

Surface tension is one of the most useful thermophysical properties that has been widely used to characterize surfaces of liquids in chemistry and chemical engineering areas such as the manufacturing of plastics, coatings, textiles and films. Surface tension is defined as the energy needed in bringing a liquid molecule to the surface for formation and maintaining the surface area in equilibrium, which is represented by surface free energy per unit surface area in thermodynamics [1,2]. Although there is a large data source of the surface tension for pure components, it is difficult to find data for the surface tension of binary systems due to the various composition and temperature ranges of liquid mixtures and the extensive amount of time needed for obtaining data by measurement using experimental devices.

The surface tension of binary systems has been predicted by using a corresponding states model proposed by Rolo et al. [4], which is based on only pure component properties such as critical volumes, critical temperatures, the acentric factors. Kahl, Wadewitz and Winkelmann have estimated surface tensions using the cubic polynomial regression in the mole fractions of binary systems [5]. Ramirez-Verduzco et al. have predicted the surface tension of binary systems as a function of concentration and temperature, which relates the surface concentration of each component to the individual activity coefficients and to the molar surface area of each component [6].

The recent studies for the prediction of the surface tension of binary systems have used mathematical models, in which procedures are more complicated than empirical models. The main objective of this study is to establish a general empirical model for the prediction of the surface tension of binary liquid systems through the regression of the reported experimental data based on the partial least squares (PLS) method. The PLS method, which was introduced theoretically by Geladi and Kowalski [3], has been applied to the prediction of the physical properties of the chemical substances such as the liquid viscosities of organic compounds by Suzuki et al. [7]. The PLS method is one of the multivariate statistical analysis methods that are empirical modeling techniques to handle the input data and the output data coincidently by using experimental data.

In this paper, the empirical model using the PLS method to predict the surface tension of unknown binary liquid systems has been developed by the correlation of thermophysical properties of only pure components and the reported experimental surface tension data of binary systems. The reported experimental data are divided into the training set for the regression and the test set for the prediction. It is illustrated that the PLS method is useful as a means of the regression of the reported experimental surface tension data and as a tool for the prediction of the surface tension of unknown binary systems.

THEORY

1. Multivariate Statistical Analysis Method

The multivariate statistical analysis method is a statistical technique by which the input-output data are analyzed simultaneously. This method is efficient for analyzing multivariate data such as complicated input-output data from industrial processes and parameters related to complex chemical reactions. It is also useful for the prediction of the product qualities and physical properties of pure components and mixtures. Multiple linear regression, principal component analysis (PCA), PLS and quadratic partial least squares (QPLS) are the representative methods of multivariate statistical analysis. The multivariate statistical analysis method is empirical modeling that relies on actual experimental data from chemical processes. The mathematical model involves various chemical reaction equations and parameters and it is difficult to make a prediction model, while the empirical model has the advantage that a prediction model could be set up easily by using the actual experimental data. In this study, the PLS method has been used in order to set up the prediction model of the surface tension of unknown binary liquid sys-

†To whom correspondence should be addressed.
E-mail: bslee@khu.ac.kr

tems through the regression of the reported experimental data with thermophysical properties of only pure substances.

2. Partial Least Squares (PLS) Method

The PLS method is a relatively new technique in multivariate statistical analysis which adopts new latent variables (LVs) to represent score vectors from input (X)-output (Y) data matrix. The latent variables are created among the input variables so as to be orthogonal with each other, highly correlated with the output variables. The LVs used in the PLS method are linearly independent of each other. There exists high correlation between the input LVs and output LVs. A model based on the PLS method consists of a regression between the LVs of the input (X) and output (Y) variable data blocks. The PLS method consists of outer relations for the input (X) and output (Y) variables data blocks and an inner relation for score vectors from the input LVs and output LVs. The input variables are adopted as latent variables in a high order of variance value.

The outer for the X and Y data blocks is, respectively,

$$X = \sum_{h=1}^{a} t_h p_h^T + E \quad (1)$$

$$Y = \sum_{h=1}^{a} u_h q_h^T + F \quad (2)$$

The input (X) variable data block is indicated by multiplication of the score vector t_h and loading vector p_h. The output (Y) variable data block is indicated by the multiplication of the score vector u_h and the loading vector q_h.

The inner relation for score vectors is

$$u_h = b_h t_h + e_h, \ (h=1, \ldots, a) \quad (3)$$

As shown in Eq. (3), the PLS method is based on a linear relationship between the input data score vectors and the output data score vectors. The PLS method is carried out in the nonlinear iterative partial least squares (NIPALS) algorithm [3] in which the above Eqs. (1)-(3) are used. The NIPALS algorithm lets the input and output data blocks have each other's scores, which gives a better inner relation between them.

PREDICTION OF THE SURFACE TENSION OF BINARY LIQUID SYSTEMS

To predict the surface tension of unknown binary liquid systems, the experimental surface tension data of various binary systems in the literature (Table 1) have been used. The reported experimental surface tension data in a training set have been regressed by using the PLS method. From the empirical model developed by the PLS method, the surface tension of unknown binary systems in the test set is predicted. In the empirical model, a total of 17 input variables having strong influences on the surface tension have been selected according to former researches [4-6]. The input data block (X) is composed of only pure component properties such as critical temperature, critical pressure, critical volume, molar volume, viscosity and vapor pressure at the various concentrations and temperatures. The reported experimental surface tension in the training set is used in the output data (Y). The prediction model of the surface tension of unknown binary liquid systems is developed while the input data block (X) and the output data (Y) are being applied to the PLS method. If a new input data block in the test set (Table 3) is applied to the prediction model, the surface tension values of unknown binary systems are calculated. For the input data block (X), 14 LVs from 17 thermophysical data of pure components are adopted for the PLS method by using cross-validation [12].

RESULTS AND DISCUSSION

To illustrate the accuracy of the PLS method, the results of the regression and the prediction as against the experimental data were compared using an average relative error (%). An average relative error for the regression results of the reported surface tension of binary systems in the training set is shown in Table 2, and the prediction results of unknown binary systems in the test set are shown in Table 3. The experimental surface tension data with the regression results of each point of binary systems are described in Figs. 1-3, prediction results in Fig. 4, respectively. Figs. 1-4 show the comparison between experimental data and results of the regression and prediction of the binary systems mentioned in Table 1. A tendency of over-prediction of the PLS method shown in Fig. 4 is not explained by the PLS method itself, but due to influences of molecular structural interaction between the components of binary system. Over-prediction or under-prediction generally occurs in the multivariate statistical analysis methods; it is recommended that the new input variables to explain the molecular interactions between the components of binary system should be introduced in further study.

Table 1. The experimental surface tension data of binary systems

	Systems	Point	Ref.
Training set	Cyclohexane+heptane	45	[5]
	Toluene+propanone	45	
	Heptane+decane	15	[4]
	Heptane+hexadecane	15	
	Hexadecane+eicosane	22	
	Benzene+nitrobenzene	18	[6]
	Pentane+butanenitrile	14	
	Iso-butanol+n-decanol	8	
Test set	Cyclohexane+propanone	36	[5]
	Decane+hexadecane	15	[4]

Table 2. The regression results of the surface tension of binary systems in the training set using the PLS method

Systems	Point	Average of relative error (%)
Cyclohexane+heptane	45	2.49
Toluene+propanone	45	1.64
Heptane+decane	15	2.35
Heptane+hexadecane	15	9.46
Hexadecane+eicosane	22	2.64
Benzene+nitrobenzene	18	1.35
Pentane+butanenitrile	14	7.25
Iso-butanol+n-decanol	8	4.37
Total	182	3.20

※ Relative errors (%) between the experiments and PLS regression

Table 3. The prediction results of the surface tension of unknown binary systems in the test set using the PLS model

Systems (test set)	Concentration		Temp. (K)	Exp. (mN/m)	PLS (mN/m)	Relative error (%)
	X1	X2				
Cyclohexane+propanone	0.1002	0.8998	287.81	23.54	23.78	1.01
	0.2001	0.7999		23.19	23.27	0.36
	0.2993	0.7007		23.23	23.40	0.73
	0.3979	0.6021		23.33	23.68	1.50
	0.4964	0.5036		23.58	24.02	1.88
	0.6000	0.4000		23.91	24.42	2.14
	0.6975	0.3025		24.31	24.82	2.08
	0.7993	0.2007		24.69	25.25	2.28
	0.8994	0.1006		25.12	25.73	2.44
	0.1002	0.8998	297.82	22.25	22.61	1.60
	0.2001	0.7999		22.02	22.22	0.90
	0.2993	0.7007		22.02	22.34	1.48
	0.3979	0.6021		22.08	22.60	2.36
	0.4964	0.5036		22.22	22.91	3.09
	0.6000	0.4000		22.53	23.25	3.21
	0.6975	0.3025		22.84	23.59	3.27
	0.7993	0.2007		23.24	23.93	2.99
	0.8994	0.1006		23.82	24.23	1.71
	0.1002	0.8998	307.86	21.03	21.27	1.13
	0.2001	0.7999		20.78	21.07	1.38
	0.2993	0.7007		20.82	21.22	1.93
	0.3979	0.6021		20.89	21.46	2.73
	0.4964	0.5036		21.14	21.73	2.79
	0.6000	0.4000		21.35	22.02	3.16
	0.6975	0.3025		21.60	22.29	3.20
	0.7993	0.2007		22.08	22.53	2.04
	0.8994	0.1006		22.58	22.58	0.01
	0.1002	0.8998	317.86	19.91	19.72	0.96
	0.2001	0.7999		19.63	19.80	0.86
	0.2993	0.7007		19.63	20.01	1.94
	0.3979	0.6021		19.70	20.25	2.79
	0.4964	0.5036		20.00	20.49	2.45
	0.6000	0.4000		20.30	20.73	2.12
	0.6975	0.3025		20.56	20.92	1.77
	0.7993	0.2007		21.04	21.03	0.03
	0.8994	0.1006		21.46	20.76	3.26
			Average of relative error (%)			1.84
Decane+hexadecane	0.250	0.750	293.15	25.29	26.71	5.61
	0.500	0.500		26.20	27.05	3.25
	0.752	0.248		27.18	27.89	2.62
	0.250	0.750	303.15	24.37	25.66	5.29
	0.500	0.500		25.48	26.07	2.33
	0.752	0.248		26.17	26.98	3.10
	0.250	0.750	313.15	23.54	24.65	4.74
	0.500	0.500		24.76	25.13	1.48
	0.752	0.248		25.46	26.09	2.47
	0.250	0.750	323.15	22.60	23.68	4.77
	0.500	0.500		23.71	24.20	2.06
	0.752	0.248		24.65	25.21	2.27
	0.250	0.750	333.15	21.67	22.71	4.82
	0.500	0.500		22.87	23.29	1.82
	0.752	0.248		23.72	24.34	2.62
			Average of relative error (%)			3.28
			Total relative error: 2.33%			

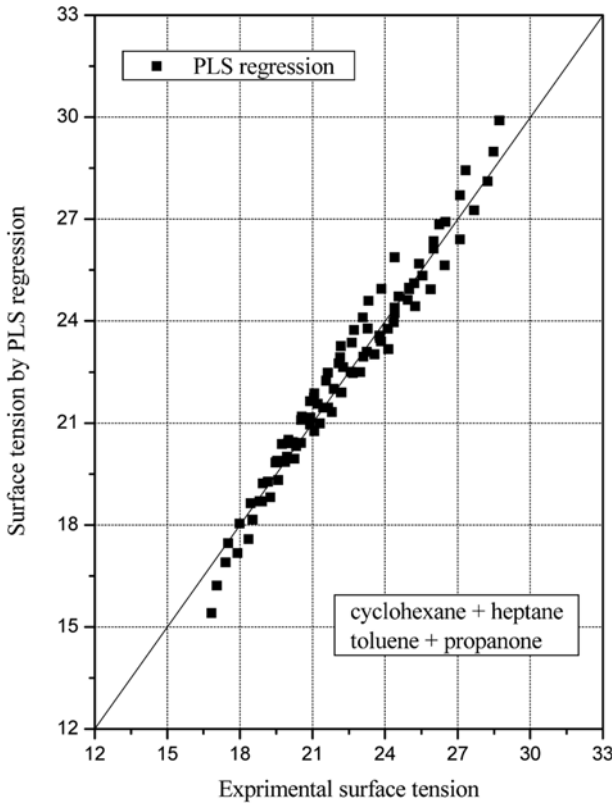

Fig. 1. The results of the regression of the surface tension of binary systems in the training set of Ref. [5].

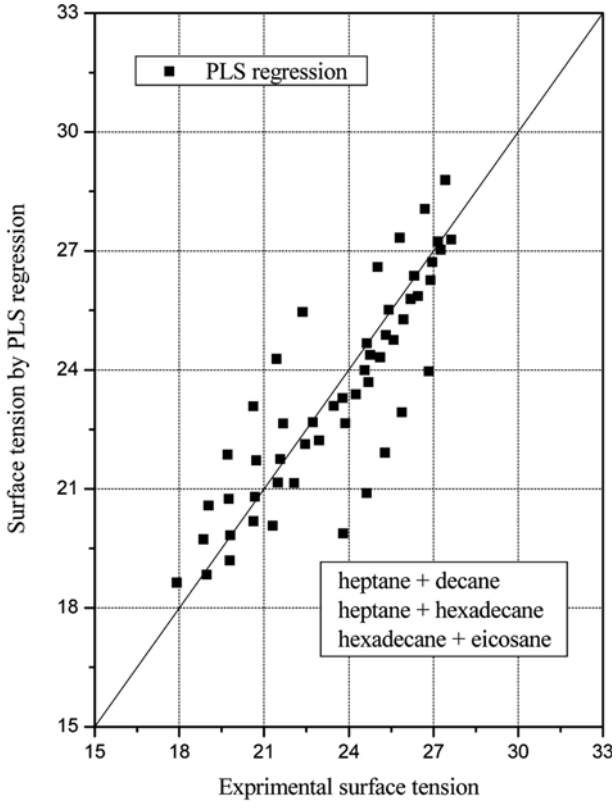

Fig. 2. The results of the regression of the surface tension of binary systems in the training set of Ref. [4].

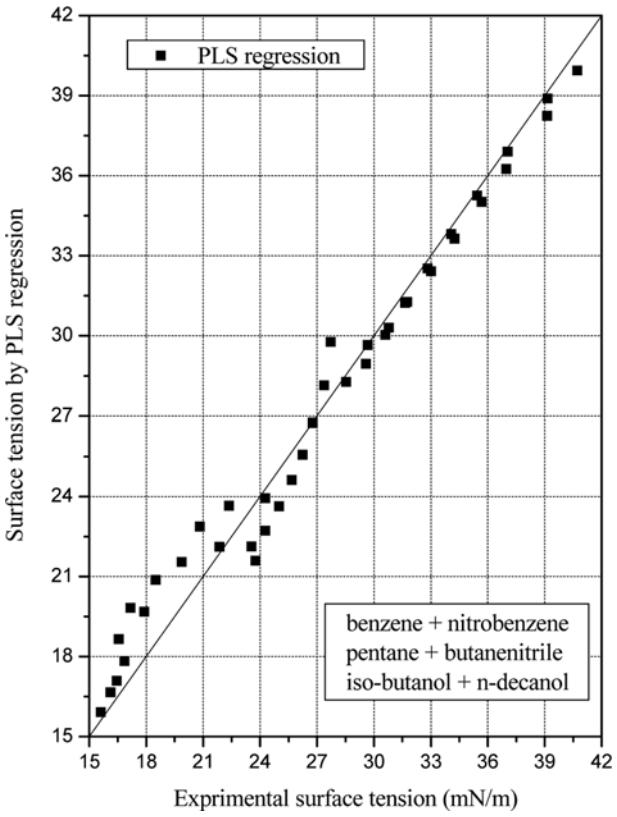

Fig. 3. The results of the regression of the surface tension of binary systems in the training set of Ref. [6].

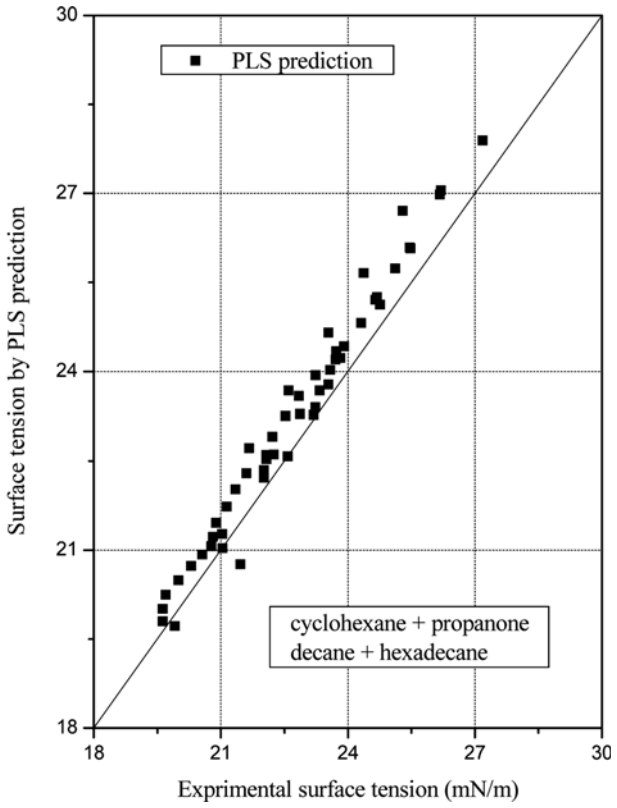

Fig. 4. The results of the prediction of the surface tension of unknown binary systems in the test set.

As shown in Table 2, an average relative error of 3.20% was obtained for the regression in 182 experimental points of a training set. Using the PLS method developed by the regression in Table 2, an average relative error of 2.33% was obtained for the prediction in 51 experimental points of a test set in Table 3. From the above results, the prediction of the surface tension of unknown binary systems using the PLS model produced outstanding results by the PLS regression of the reported experimental data and only pure component properties. The binary systems which include cyclohexane+propanone and decane+hexadecane showed the relative errors (%) of 1.84% and 3.28%, respectively, in the prediction model. The mathematical model [4,5] based on the corresponding states model and the cubic polynomial regression has indicated relative errors of 5% for the specified mixtures. It is possible that the surface tension of unknown binary systems composed of pure components not in the training set can be predicted; however, the accuracy of the prediction is unsatisfactory in this case. According to this study, unknown binary systems composed of pure components in the training set give outstanding prediction results. As a result of the above consideration, the prediction of the surface tension by the PLS method is superior to the other mathematical models. To further improve this method, we highly recommend that the information of the molecular structure and the other thermophysical properties of pure components must be considered in the input data block (X) of the PLS model.

CONCLUSION

This paper has suggested the PLS method for the regression of the reported surface tension of binary systems and the prediction of the surface tension of unknown binary systems based on the reported experimental data and only pure component properties such as critical temperature, critical pressure, critical volume, molar volume, viscosity and vapor pressure. With the PLS method, the surface tension of unknown binary liquid systems can be predicted by the empirical model derived from the regression of the reported binary systems. The PLS method for the prediction of the surface tension of binary systems is superior to the other mathematical models. It is easy to use and the accuracy may be satisfactory for predicting the surface tension of unknown binary liquid systems.

NOMENCLATURE

X : input data block
Y : output data block
t_h : score vector from X
u_h : score vector from Y
p_h^T : loading vector from X
q_h^T : loading vector from Y
E : residual from $E = X - \sum t_h p_h^T$
F : residual from $F = Y - \sum u_h q_h^T$
b_h : the regression coefficients between the score vectors from the X and Y

REFERENCES

1. C. A. Miller and P. Neogi, *Interfacial phenomena: equilibrium and dynamic effects*, Marcel Dekker, Inc., New York (1985).
2. J. Lykema, G. J. Fleer, J. M. Kleijn, F. A. M. Leermakers, W. Norde and T. Van Vliet, *Fundamentals of interface and colloid science, volume III: Liquid-fluid interfaces*, Academic Press (2000).
3. P. Geladi and B. R. Kowalski, *Analytica Chemica Acta.*, **185**, 1 (1986).
4. L. I. Rolo, A. I. Caco, A. J. Queimada, I. M. Marrucho and J. A. P. Coutinho, *J. Chem. Eng. Data*, **47,** 1442 (2002).
5. H. Kahl, T. Wadewitz and J. Winkelmann, *J. Chem. Eng. Data*, **48**, 1500 (2003).
6. L. F. Ramirez-Verduzco, A. Romero-Martinez and A. Trejo, *Fluid Phase Equilibria*, **246**, 119 (2006).
7. T. Suzuki, K. Ohtaguchi and K. Koide, *Computers & Chemical Engineering*, **20**(2), 161 (1996).
8. S. Y. Kim, B. Lee, C. B. Chung and S. H. Choi, *Journal of the Korean Institute for Gas*, **10**(4), 29 (2006).
9. J. Aguil-Hernandez, I. Hernandez and A. Trejo, *International Journal of Thermophysics*, **16**(1), 45 (1995).
10. H. Kahl, T. Wadewitz and J. Winkelmann, *J. Chem. Eng. Data*, **48**, 580 (2003).
11. A. J. Queimada, I. M. Marrucho and J. A. P. Coutinho, *Fluid Phase Equilibria*, **183-184**, 229 (2001).
12. P. Bastien, V. E. Vinzi and M. Tenehaus, *Computational Statistics & Data Analysis*, **48**, 17 (2005).
13. A. J. Burnham, J. F. MacGregor and R. Viveros, *Chemometrics and Intelligent Laboratory Systems*, **48**, 167 (1999).
14. S. Wold, J. Trygg, A. Berglund and H. Antti, *Chemometrics and Intelligent Laboratory Systems*, **58**, 131 (2001).