Presented at the 3rd Chemical Engineering Conference for Collaborative Research in Eastern Mediterranean, May 13 –15, 2003, Thessaloniki, Greece

A DYNAMIC LIBRARY FOR PHYSICAL AND THERMODYNAMIC PROPERTIES CORRELATIONS

Tatyana Spivak, Chem. Eng. Dept., Ben-Gurion University, Beer-Sheva 84105, Israel;
Mordechai Shacham[1], Chem. Eng. Dept., Ben-Gurion University, Beer-Sheva 84105, Israel;
Neima Brauner, School of Engineering, Tel-Aviv University, Tel-Aviv 699 78, Israel;

Abstract

Analysis of the correlations included in the existing thermodynamic property libraries was carried out in order to identify the aspects that need to be improved. In those libraries no information is provided regarding the accuracy and stability of the correlations included. It has been demonstrated that those correlations are often unstable, meaning that their asymptotic behavior and/or extrapolation even in a small range can be completely erroneous, and addition or removal of data points can cause dramatic change in the parameter values. In other cases the correlations are much less accurate than the achievable accuracy based on the precision of the data. To overcome these limitations, we have developed a new, prototype library, where some of the deficiencies are corrected. Advanced stepwise regression techniques (the SROV program, Shacham and Brauner, 2003) are used to fit the most accurate and stable model to the data. In addition to the correlations and the experimental data, the library contains information regarding the quality and precision of the correlation: residual plots, confidence intervals on the parameter values and the variance. The library can be found in the web-site http://www.bgu.ac.il/chem_eng/thermo_lib/thermo_lib.htm

Keywords

Correlation library, stepwise regression, stability and precision of correlations, process modeling

Introduction

Correlation equations of physical and thermodynamic properties are being used extensively in process calculations and computations. Those correlations provide the property values as function of process conditions, such as temperature, pressure and composition. The correlations are usually developed by fitting regression models (equations) to experimental data (or using estimated data when experimental data are not available). They are collected to form libraries of physical and thermodynamic properties. The libraries contain correlations for many different properties for a large number of chemical compounds. One of the first, better known libraries were published in the book of Reid *et al* (1977). Libraries that include more property correlations for many more components are available now (see, for example, Daubert and Danner, 1998).

Development of new, more accurate and stable regression models for various properties is an ongoing continuous process. Such development is made possible by availability of more accurate experimental data, development of new stepwise regression and regression diagnostic techniques (Shacham and Brauner, 1999 and 2003) and availability of high speed high precision computers, where the regression model complexity becomes a secondary concern to the accuracy of the calculated properties.

Unfortunately, the current correlation libraries cannot be regularly updated to include the most accurate and stable correlations for the various properties. The structure of those libraries is static, in the sense that there is one form of correlation for a particular property and the library cannot accept a new correlation in a different form, even if it is much more accurate. The libraries rely on old data, which is often not real experimental data, but smoothed,

---

[1] To whom all correspondence should be addressed

interpolated (often even extrapolated) data, the precision of which is very difficult to assess.

In this paper some of the limitations of the existing physical properties libraries are demonstrated. The Web based version of the DIPPR library (http://dippr.byu.edu) is used for this demonstration. A new type of web based library (http://www.bgu.ac.il/chem_eng/thermo_lib/thermo_lib.htm) that we have developed, which overcomes the limitations of the existing libraries, is described. The correlations included in this library were developed using the SROV stepwise regression program (Shacham and Brauner, 1999) for multi-linear and polynomial regressions and the POLYMATH[2] software package for nonlinear regression.

Some examples of the limitations of the existing libraries

One of the most severe limitations of the existing correlation libraries stems from the need to use the same type of correlation for a particular property, regardless of the range, precision and trend of the data available. This may result in inclusion of too many explanatory variables (the independent variable and its functions) in the correlation, which renders an unstable model. Typical ill effects characterize the instability, whereby adding or removing an experimental point from the data set may drastically change the parameter values, and extrapolation even in a very short range may yield absurd results. An insufficient number of explanatory variables, on the other hand, may result in an inaccurate model, where the residual plot does not show a random error distribution, but clear trend with some very large residuals. These drawbacks will be demonstrated using some correlations and data for ethane from the DIPPR library.

Redundant free parameter in the regression model : For correlation of the solid heat capacity of ethane, DIPPR recommends using a fourth order polynomial of the form $Cp = A + BT + CT^2 + DT^3 + ET^4$. Considering the value of the associated linear correlation coefficient ($R^2 = 0.9998$), the correlation seems to fit excellently the experimental data. However, the ratio between the value of the free parameter A (= 8158.44) and its 95% confidence interval (=12780) indicates that the free parameter is probably redundant and may possibly be removed from the correlation. Further enforcement of this conclusion is obtained by looking at Figure 1, where the calculated curve and the experimental data for solid heat capacity are plotted. It can be seen that when the temperature approaches the absolute zero (0 K), the heat capacity increases to the value of 8158 instead of approaching zero, as expected from theoretical considerations. Removing the free parameter from the correlation does not affect noticeably its accuracy ($R^2 = 0.9997$), but renders the right asymptotic behavior when approaching 0 K, and all the 95% confidence intervals are smaller than the respective parameter values.

Too many explanatory terms in the correlation: For correlation of vapor pressure of solid ethane, DIPPR recommends the use of the truncated Riedel's equation: $\ln P = A + B/T + C \ln T$, where P is vapor pressure in Pa. The coefficient values (including 95% confidence intervals) obtained with POLYMATH when using the four data points recommended by DIPPR, are $A = -188.64 \pm 93.17$, $B = 923.15 \pm 1455.9$ and $C = 39.65 \pm 17.11$ with $R^2 = 1.0$. While the correlation coefficient indicates a perfect fit, the wide confidence interval on the parameter B may point to "over-correlation" with too many explanatory variables included in the regression model. Indeed, when an additional, fifth data point is added to the data set (this data point is also listed in DIPPR, but was taken from a different source than the first four data points) the following, completely different parameter values are obtained: $A = -582.92 \pm 2189.9$, $B = 7076.3 \pm 34230$ and $C = 112.09 \pm 402.24$ with $R^2 = 0.995$. Thus, this correlation is extremely unstable. However, it can be simply stabilized, without a significant reduction of the accuracy, by removing the $\ln T$ term from the equation. The parameter values obtained in this case are: $A = 27.2387 \pm 1.6627$, $B = -2449.86 \pm 140.77$ and $R^2 = 0.9996$ with the four data points, and $A = 27.3453 \pm 4.9967$, $B = -2462.44 \pm 422.74$ and $R^2 = 0.9913$ when the five data points are used.

Insufficient number of parameters in the model: For correlation of solid heat capacity of ethane DIPPR recommends the use of a four-parameter correlation.

---

[2] POLYMATH is copyrighted by M. Shacham, M. B. Cutlip and M. Elly (http://www.polymath-software.com)

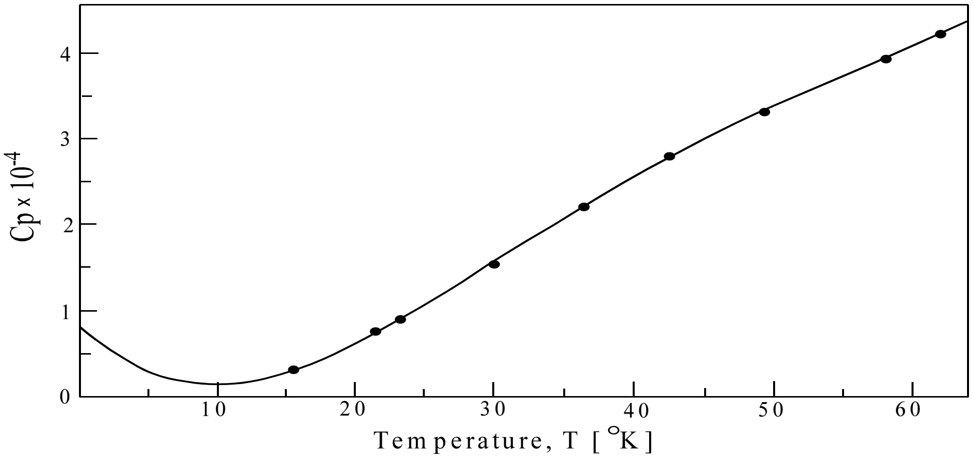

Figure 1: Calculated curve and experimental data for heat capacity of solid ethane (with a free parameter)

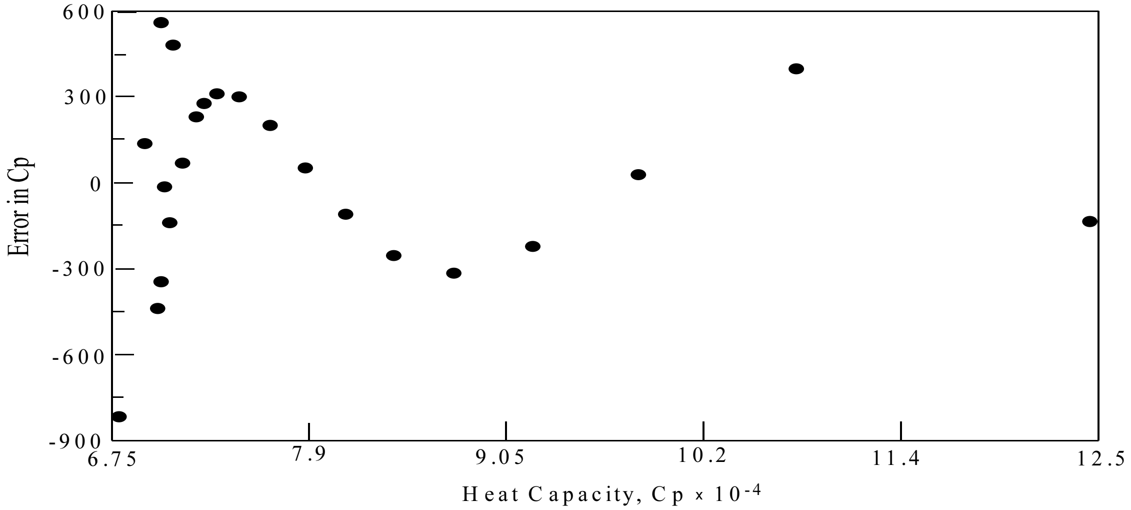

Figure 2: Residual plot of the DIPPR library correlation of liquid heat capacity data

The residual plot of this correlation is shown in Figure 2. It can be seen that the distribution of the residuals exhibits a cyclic pattern, where the maximal absolute error is about 800. The optimal correlation identified by SROV for this property is a polynomial model with seven parameters, which includes the polynomial terms z, z^2, z^4, z^5, z^{10} and z^{13}, where z is the standardized temperature $z = (T - 185.7391)/62.1795$. The residual plot for this correlation exhibits a random residual distribution with a maximal error less than 30. Thus, using the appropriate model, a correlation of a much higher accuracy is obtained, whereby the maximal error is reduced by more than an order of magnitude.

Structure of the proposed library

The basic unit of the proposed library is the compound – property worksheet. In the prototype library there are about 270 such worksheets, for different types of compounds and different properties. A compound-property worksheet contains the following information: name of the compound, name of the property, literature reference for the data, temperature range of the data, reported experimental error in the temperature and property data, the experimental data as reported in the reference, the optimal correlation obtained, plot of the experimental data points and the calculated curve versus temperature (or standardized temperature) and the residual plot.

In order to obtain accurate, stable and reliable correlations, only high precision and consistent experimental data are used. Those were published in the Journal of Chemical and Engineering Data and Fluid Phase Equilibria in recent years. First we have attempted to fit the data using the correlations recommended by the DIPPR (Daubert, 1998). If those correlations did not yield satisfactory fit (as indicated by non-randomness of the residual plot), we have attempted fitting other types of models, using the SROV program, until a stable model with random residual distribution was obtained.

Polynomial models were used to fit heat capacity, thermal conductivity, density and kinematic viscosity data. The SROV program was used to select the most appropriate, non-consecutive polynomial terms that should be included in the correlation. At first, fitting the data in polynomial terms of the temperature was attempted. If a need to include high powers of temperature arose (such as above T^5), we switched to the use of standardized temperature values: $z = [T\text{- mean }(T)]/\text{stdev}(T)$, to avoid the numerical ill-conditioning introduced by the huge numbers of the high powers of the temperature. The SROV program uses the reported experimental error, and stops adding polynomial terms to the correlation when it diagnoses an insignificant signal-to-noise ratio in the remaining terms. In such cases, adding more terms actually results in modeling of the experimental error. It is worth noting that if the reported experimental error is larger than the actual error in the data, the SROV stops adding terms prematurely. Such a situation is also indicated by a non-random residual distribution. In such cases we inputted a smaller experimental error estimate to SROV than reported in the literature source of the data. In a few cases there was also a need to carry out Box – Cox (maximum likelihood) transformation of the measured property in order to ‘linearize’ the regression equation. Consequently, in some cases the logarithm of the property was modeled instead of the property itself.

For vapor pressure, in most cases, various powers of $\tau = 1\text{-}T/Tc$ were selected to model $\ln(P/Pc)$, as recommended by Wagner (1973), using the SROV program. In other cases polynomial terms were added to the Clapeyron or the Riedel's equations to compensate for the deviation from the straight line of the logarithm of the vapor pressure versus the reciprocal of the temperature. This new type of model is introduced here for the first time. The same type of correlation was used mostly for modeling also viscosity data.

In all the cases tested, stable correlations of minimum variance were identified. These are associated with random residual distributions and statistically significant parameters.

Conclusions

It has been shown that the correlations presented in the existing thermodynamic property libraries are often unstable, meaning that their asymptotic behavior and/or extrapolation even in a small range can be completely erroneous, and addition or removal of data points can cause dramatic changes in the parameter values. In other cases the correlations are much less accurate than the achievable accuracy based on the precision of the data. The user of the library is not provided with any information regarding the accuracy and stability of the proposed correlation.

In the new type of library, a prototype of which we have developed, most advanced stepwise regression techniques are used to fit the most accurate and stable model to the data. The library also contains information regarding the quality and precision of the correlations in forms of residual plots, confidence intervals on the parameter values and the variance.

The use of this new type of correlation library will help in developing more accurate and reliable simulation models for chemical processes.

References

Daubert, T.E. and Danner, R.P., *Physical and Thermodynamic Properties of Pure Chemicals: Data Compilation*, Hemisphere Publishing Co., New York (1989).

Daubert, T. E., *Ind. Eng. Chem. Res*. **37**, 3260, (1998)

Reid, R. C., Prausnitz, J.M. and Sherwood, T.K., *The Properties of Gases and Liquids,* 3rd Ed, McGraw-Hill, New York (1977).

Shacham, M. and N. Brauner, *Chem. Eng. Process*. **38**, 477,(1999)

Shacham, M. and N. Brauner, *Comput. Chem. Engng*. (in press, 2003)

Wagner, W., Cryogenics 13, 470, (1973)