CORRELATIONS

Modeling Physical and Thermodynamic Properties via Inverse Normalizing Transformations

Haim Shore[†]

Department of Industrial Engineering and Management, Ben Gurion University of the Negev, Beer-Sheva 84105, Israel

Neima Brauner[‡]

School of Engineering, Tel-Aviv University, Tel-Aviv 699 78, Israel

Mordechai Shacham*

Department of Chemical Engineering, Ben Gurion University of the Negev, Beer-Sheva 84105, Israel

The inverse normalizing transformation (INT) represents a generalization of the inverse of the Box−Cox transformation. It is shown that several well-known and widely used property correlation equations, such as the Antoine, the truncated Riedel, the Rackett, and the Guggenheim equations can be derived from the INT. Its use is demonstrated for modeling the temperature dependence of vapor pressure, solid and liquid heat capacity, vapor and liquid viscosity, and surface tension data. It is shown that the INT can represent satisfactorily curves of different shapes and, as such, its use can be beneficial in modeling the temperature dependence of various physical and thermodynamic properties.

1. Introduction

Mathematical modeling and simulation of chemical processes and substances require precise equations to represent the physical and thermodynamic properties as a function of state variables such as temperature, pressure, and composition. Most of the model equations that are used for calculating or predicting these properties are based on a physical theory,[1] while the theory-based terms are often complemented by empirical terms in order to achieve higher accuracy. The parameters of the model are evaluated by regression analysis of experimental data. If the latter appear to be scarce, it is often necessary to use the fitted expression for extrapolation to regions where experimental data are not available.

There are many physical theory-based relationships that can represent accurately data for a certain class of substances and for a particular property. Fitting procedures that would help determine the parameters of these relationships are routinely based on the least-squares methodology, and if linearization of the relationship is feasible, some form of a stepwise linear regression is commonly employed (refer, for example, to ref 2). In the latter case, ensuring the validity of the underlying assumptions of linear regression analysis (like the normality of the residuals) is essential to guarantee the validity of the derived conclusions.

There are, however, many properties (like solid properties) where the systematic variation of the inves-

tigated property as a function of some state variables is not understood well enough to construct theory-based model equations. Furthermore, even if a theoretically based first approximation does exist, discrepancies are sometimes observed that still require an empirical refining of the approximation. In such cases, some general empirical equations (polynomials, for example) are often used. Polynomials, however, may be highly inaccurate unless high-degree powers are used. In the latter case, unwarranted inflections may be introduced which render interpolation and extrapolation very inaccurate. Furthermore, attempting to use unduly high degree polynomials for noisy data may result in "over-fitting" that could cause a high level of collinearity between the polynomial terms. This may render the polynomial representation highly unstable (see, for example, ref 3).

A third situation where empirical equations may be needed is when the accurate theory-based models may require critical constants and/or a large set of data for obtaining reliable estimates of the model parameters. Often, not enough data are available, and attempting to use models that contain many parameters when data are insufficient may lead to poor or even misleading and absurd results.

In this paper we apply a new statistical-distribution theory-based equation to model the temperature dependence of physical and thermodynamic properties. The general form of the equation has been developed by Shore,[4] and it may be considered to be a generalization of the inverse of the well-known Box−Cox[5] normalizing transformation, which has been widely applied since its inception in 1964. The equation proposed by Shore[4] models the quantile of a dependent variable, which is not necessarily normal, in terms of the corresponding standard normal quantile. The resulting expression is

---

* To whom correspondence should be addressed. Phone: 972-818-6461481. Fax: 972-818-6472916. E-mail: shacham@bgumail.bgu.ac.il.

[†] E-mail: shor@bgumail.bgu.ac.il.

[‡] E-mail: brauner@eng.tau.ac.il.

called an inverse normalizing transformation (INT), and the normal and the log-normal quantile functions are exact special cases. Several derivatives of the "origin" INT, with a reduced number of parameters, have also been developed. Their accuracy was demonstrated to many and variously shaped distributions like the binomial, Poisson, gamma, Weibull, and extreme value. The theory underlying the new INTs seems to be as valid and applicable when the relationship between various physical and thermodynamic properties is the subject of the modeling effort.

The structure of the paper is as follows. In the next section (section 2), some basic concepts including the Box−Cox[5] transformation are briefly reviewed. In section 3, we outline the theory behind the new group of INTs and depict some of the new INTs. In section 4, the relationship between the general INT equation and some physical theory-based property correlation equations is investigated. In section 5, a particular form of the general equation (the simplified INT, SINT), found to be most appropriate for physical and thermodynamic property correlation, is presented. The SINT is tested and compared with existing correlation equations by correlating vapor pressure, solid and liquid heat capacity, vapor and liquid viscosity, and surface tension data, in section 6. Section 7 summarizes the study and offers some conclusions.

The calculations reported in this study were carried out with Mathematica [a trademark of Wolfram Research, Inc. (http://www.wolfram.com)] and POLYMATH 5.0 [copyrighted by M. Shacham, M. B. Cutlip, and M. Elly (http://www.polymath-software-.com)] computational packages.

2. Basic Concepts

Let us assume that there is known to be a relationship between a dependent variable Y and an independent variable X. Assume further that there is a set of N data points, $\{x_i, y_i\}$, which may be used to model and estimate this relationship. A general form of the regression model to represent Y in terms of X is

$$y_i = F(x_i, \beta) + \epsilon_i \tag{1}$$

where β is a vector of p parameters and $\{\epsilon_i\}$ are independent normally distributed errors with constant variance. The values of the parameters $\beta_0, \beta_1, ..., \beta_p$ are usually estimated to minimize the sum of squares of the errors:

$$s^2 = \min_{\beta} \sum_{i=1}^{N} [y_i - F(x_i, \beta)]^2 \tag{2}$$

Dependent on the nature of function F, parameter fitting employed by the least-squares approach is carried out by either linear or nonlinear regression techniques. Various models fitted to the same data are usually compared using the value of the variance, the distribution of the residuals, and the confidence intervals associated with the estimated parameter values. A smaller variance indicates a better fit, while a nonrandom residual distribution may imply an inappropriate model. Large confidence intervals (in particular, confidence intervals that are larger than the respective parameter values) may indicate the presence of collinearity among the terms containing the independent variable X (overcorrelation) or the presence of outliers in the data. For a general discussion of regression diagnostics, see, for example, the book by Belsley et al.[6]

For application of regression diagnostics to the modeling of thermophysical data, refer to refs 7 and 8.

A large variety of regression models have been proposed for representing the temperature dependence of various physical and thermodynamic properties. The most widely used models have been recently reviewed and compared by Daubert.[9] Most model equations are based on physical theory (Reid et al.[1]), while some empirical terms are added to compensate for variations not accounted for by the existing theory. Addition of empirical terms is done by use of stepwise regression (see, for example, ref 2) or by simple trial and error.

A different approach to modeling, which has gained popularity also in chemical engineering applications, is to transform the dependent variable so that when a linear model for F is used, constancy of the error variance and normality of the error distribution are obtained. Most widely used in this context is the Box−Cox[5] transformation. The one-parameter version of the latter is

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda} & (\lambda \neq 0) \\ \log(y) & (\lambda = 0) \end{cases} \tag{3}$$

If this transformation is effective, then $y^{(\lambda)}$ is expected to be well represented by a linear function of a standard normal variable x, $F(x, \beta) = \beta_0 + \beta_1 x$. The value of λ that best satisfies this condition can be found by the "maximum likelihood" procedure, described in ref 5. The maximum likelihood method was implemented, for example, in the Simusolv program,[10] which has been widely used for reaction-rate data correlation. It has been noted[11] that the Box−Cox transformation is useful in the correction of skewness of the distribution of error terms, unequal error variances, and nonlinearity of the regression function. Thus, this transformation can be helpful in modeling the temperature dependence of physical and thermodynamic properties, in particular, in cases where the change of the property as a function of the temperature is not well enough understood to enable the development of a physical theory-based model equation.

In the next section, the INT method, which stems from a generalized representation of the inverse of the Box−Cox transformation, is presented. This method would provide a general framework for modeling the regression equations derived in this paper.

3. General Form of an INT (Shore[4])

The Box−Cox transformation can be used as the basis for developing an INT from which functions representing curves of various shapes and having various error distributions can be derived. Based on the premise that $y^{(\lambda)}$ can be represented as a linear function of x, eq 3 yields

$$y = \begin{cases} [\lambda(\beta'_0 + \beta'_1 x) + 1]^{1/\lambda} = M(1 + \beta_0 x)^{\beta_1}, & \lambda \neq 0 \\ \exp(\beta_0' + \beta_1' x) = M \exp(\beta_2 x), & \lambda = 0 \end{cases} \tag{4}$$

where the middle term in each case is rewritten on the right-hand side to preserve the median of Y, M.

Shore[4] examined a large number of differently shaped distributions and found eq 4 to deliver generally poor representation of the quantile relationship between y (the quantile of a nonnormal variable Y) and the corresponding standard normal quantile, X. To improve

the fit, let us rewrite the two parts of eq 4 as a product of the two terms, thus effectively eliminating the discontinuity built into the original Box−Cox transformation:

$$y = M\{\exp[\beta_1 \log(1 + \beta_0 x) + \beta_2 x]\} \quad (5)$$

Generalizing the log term by presenting it as a Box−Cox transformation (therefore, introducing an additional parameter), we obtain from eq 5

$$y = M\exp\{\beta_1[(1 + \beta_0 x)^{\beta_3/\beta_0} - 1]/$$
$$(\beta_3/\beta_0) + \beta_2 x\} \quad x > -1/\beta_0 \quad (6)$$

This is a four-parameter INT, where the median serves as a scale parameter and is preserved by the INT (namely, when $x = 0$, $y = M$). The remaining four parameters need to be determined by an appropriate fitting procedure. Note that the normal ($\beta_2 = 0$, $\beta_3 = 0$, and $\beta_1 = 1$) and the log-normal ($\beta_3/\beta_0 = 1$) are exact special cases of eq 6. It is easy to see that, by proper choice of the parameters, different functions which are widely used to represent monotonically convex (or concave) functions can be derived. A comprehensive description of these functions may be found in ref 4. In particular, for $\beta_2 = 0$ and $\beta_3 = 0$, the equations $y = A + Bx$ (linear increase, $\beta_1 = 1$) and $y = (A + Bx)^k$ (power increase, $\beta_1 \neq 1$), where k is a real-valued parameter, are obtained (these are the cases derived from the inverse of the Box−Cox transformation). Introducing $\beta_3/\beta_0 = 1$ yields an "exponential increase" function, $y = \exp(A + Bx)$, while for $\beta_0 \rightarrow 0$, the "exponential−exponential increase" function is obtained.

4. Deriving Physical Theory-Based Models from the INT Equation

The examples shown at the end of the previous section demonstrate the versatility of the new INT and its potential to serve for modeling various relationships that are commonly encountered in chemical engineering. In particular, it may be used to represent the change of physical and thermodynamic properties as a function of temperature. This option will be explored by deriving the Antoine[12] equation for vapor-pressure correlation, Rackett's[13] equation for liquid density correlation, Guggenheim's[14] equation for surface tension correlation, and a truncated Riedel's equation for liquid viscosity correlation.

The Antoine equation is usually written as

$$\ln P = A + \frac{B}{T + C} \quad (7)$$

where A, B, and C are constants determined by the regression of experimental data. This equation is widely used for vapor-pressure correlation and gives a fairly accurate representation provided that the temperature range is not too wide. The Antoine equation can be derived from eq 6 by introducing $y = P$, $x = T$, $\beta_3/\beta_0 = \lambda = -1$, $\beta_2 = 0$, $1/\beta_0 = C$, $\beta_1/\beta_0 = B$, and $\ln(M) + \beta_1 = A$, resulting in $\beta_0 = -\beta_3 = 1/C$, $\beta_1 = B/C$, $\beta_2 = 0$, and $\ln(M) = A - B/C$.

Rackett[13] has suggested the following equation for liquid density in terms of temperature:

$$\log(\rho_L) = \log(A) - \left[1 + \left(1 - \frac{T}{C}\right)^D\right]\log(B) \quad (8)$$

where ρ_L is the liquid density and A, B, C, and D are

parameters to be determined. Daubert[9] compares Rackett's formula with other alternatives that have appeared in the literature and concludes that Rackett's equation is the most appropriate for modeling the liquid density over the entire range of temperatures from the triple point to the critical point. When eq 8 is compared with eq 6, it is obvious that eq 8 is a special case of eq 6, with $y = \rho_L$, $x = T$, $\beta_0 = -1/C$, $\beta_1 = -D \log B$, $\beta_2 = 0$, $\beta_3 = -D/C$, and $M = -A/B^2$.

Guggenheim[14] has suggested the following correlation equation for surface tension:

$$\sigma = \alpha\left[1 - \left(\frac{T}{T_C}\right)^C\right] \quad (9)$$

After comparing several correlation equations, Daubert[9] recommends the use of Guggenheim's equation for correlating surface tension data. It can be shown that eq 9 is a linear transformation of a special case of eq 6. Introducing in the latter $y = \sigma$, $1 + \beta_0 x = T/T_C$, $\beta_3/\beta_0 = \lambda = 0$, $\beta_1 = C$, $\beta_2 = 0$, and $M = \alpha$, we obtain

$$M\exp\left\{(\beta_1/\lambda)\left[\left(\frac{T}{T_C}\right)^\lambda - 1\right]\right\} = \alpha\exp\left[\log\left(\frac{T}{T_C}\right)^C\right] = \alpha\left(\frac{T}{T_C}\right)^C \quad (10)$$

which is a linear transformation of eq 9.

For viscosity of liquids, Daubert[9] recommends the use of a truncated Riedel[15] equation:

$$\log(\mu_L) = A + \frac{B}{T} + C\log(T) \quad (11)$$

It is claimed therein that "an evaluation with two dozen disparate compounds with data available over most of the range for most compounds confirmed the accuracy of this expression in correlating, interpolating, and extrapolating liquid viscosity data".

It is can be seen from eq 6 that with $\beta_3/\beta_0 = \lambda = -1$ the reciprocal term in eq 11 is obtained, while with $\lambda = 0$ the log term is obtained. Thus, instead of linearly *combining* these two terms, it is expected that an intermediate value of λ will provide a good fit for the liquid viscosity that would be at least as good as that delivered by eq 11.

5. SINT

Inspection of the curves representing the temperature dependence of various physical and thermodynamic properties (see, for example, ref 10) shows that those functions are monotonically convex (concave) functions of temperature. Assuming that the temperature (the independent variable) has a symmetric distribution, then the property (say y) is positively skewed (negatively skewed). From a comprehensive study of theoretical distributions,[4] with a skewness value that ranged from zero to over 11, β_0 was consistently found to be very small. Assuming $\beta_0 \rightarrow 0$, eq 6 can be transformed into the following double-exponent form:

$$y = a\{\exp(b[\exp(c\theta) - 1] + d\theta)\} \quad (12)$$

where θ is the standardized temperature: $\theta = [T - \text{median}(T)]/[\text{s.d.}(T)]$, T is the temperature in K, y is the pertinent physical property, and a, b, c, and d are the model parameters. Note that β_1, β_2, and β_3 are renamed as b, c, and d for the sake of simplicity and the median of y, M, is replaced by an additional adjustable param-

Table 1. Data Characterization and Fit of the SINT Equation: Vapor-Pressure Data

	oxygen		argon		nitrogen	
parameter	value	confidence interval	value	confidence interval	value	confidence interval
$a' = \ln(a)$	0.305 958 7	9.79×10^{-5}	2.502 048	2.73×10^{-4}	1.841 746	5.08×10^{-4}
b	−0.855 077	0.004 937 5	−0.857 77	0.009 228 7	−0.963 51	0.012 837 1
c	−0.603 461	0.001 565 1	−0.718 57	0.003 376 3	−0.705 34	0.004 006 1
d	0.533 671 2	0.001 530 1	0.626 196	0.003 415 9	0.626 073	0.004 672 1
variance	1.76×10^{-7}		4.33×10^{-7}		2.04×10^{-6}	
sample size (N)	183		57		68	
temp units	K		K		K	
temp range	90.188	154.581	83.804	150.651	63.148	126.2
median T	125.1		120.1		97.04	
pressure units	MPa		bar		bar	
pressure range	0.101 28	5.043 37	0.6895	48.578	0.1252	34.002
median $\ln(P)$	0.3008		2.5047		1.84	

eter a. Also note that, because T is assumed to be symmetrically distributed, standardization is carried out with the median of T rather than with the conventional mean.

Equation 12 can be rewritten in a form commonly used in the correlation of various properties by taking the logarithm of both sides:

$$\ln(y) = a' + b[\exp(c\theta) - 1] + d\theta \qquad (13)$$

where $a' = \ln(a)$. This form contains only one parameter in a nonlinear expression and thus can alleviate the computational complexity in deriving the least-squares estimates of the parameter values. Equations 12 and 13 will, henceforth, be denoted as simplified INT (SINT). This SINT will be used for the correlation of vapor pressure, heat capacity of solid and liquid, viscosity of gas and liquid, and surface tension of liquid. In the following section, the details of the performance of these analyses are described.

6. Correlation of Physical and Thermodynamic Properties with the SINT

Vapor Pressure between the Triple Point and the Critical Point. Wagner et al.[2,16] had published high-precision data for the vapor pressure for argon, nitrogen, and oxygen. The ranges of the data were from the triple point to the critical point for argon and nitrogen and from the normal boiling point to the critical point for oxygen. Some additional information for characterization of these data, together with the results of fitting eq 13 to these data, is given in Table 1.

It can be seen that the regressed model fits the data very well. The variance [based on $\ln(P)$] ranges from 1.76×10^{-7} for oxygen up to 2.04×10^{-6} for nitrogen. All of the confidence intervals on the model parameters are smaller by, at least, 2 orders of magnitude than the parameter values. The experimental vapor-pressure data and calculated curve (for oxygen) are plotted versus the temperature in Figure 1, showing that all of the experimental data points are virtually located on the calculated curve (very small errors). Similar results were obtained for argon and nitrogen. It should be noted that the vapor pressure increases exponentially with the temperature and the curve has a convex shape.

The performance of SINT for a wide range of vapor-pressure data was compared with the performance of the Antoine equation (eq 7), Wagner's[2] equation, and the extended Riedel equation.[9] The Wagner equation is the four-parameter equation:

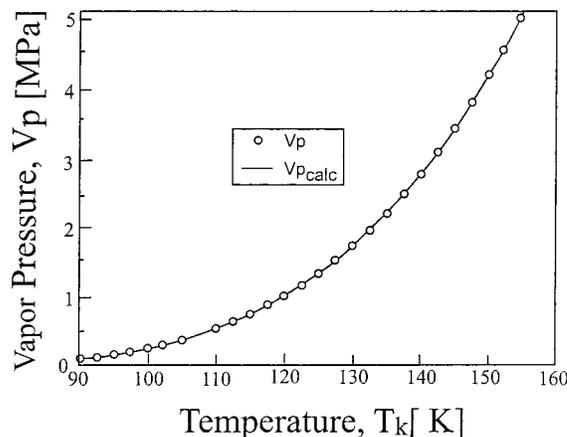$$\ln P_R = \frac{a\tau + b\tau^{1.5} + c\tau^3 + d\tau^6}{T_R} \qquad (14)$$

Figure 1. Experimental data and calculated curve (using SINT) for oxygen's vapor pressure.

Table 2. Comparison of Vapor-Pressure Correlation Equations

correlation eq (no. of param)	variance at optimal solution		
	oxygen	argon	nitrogen
Wagner's (4)	6.32×10^{-9}	4.62×10^{-8}	4.726×10^{-8}
modified Riedel's (5)	1.38×10^{-7}	1.83×10^{-7}	2.73×10^{-7}
SINT (4)	1.76×10^{-7}	4.33×10^{-7}	2.04×10^{-6}
Antoine (3)	4.06×10^{-5}	5.09×10^{-5}	7.85×10^{-5}

where a, b, c, and d are adjustable parameters, T_R is the reduced temperature ($T_R = T/T_C$, where T_C is the critical temperature of the particular substance), P_R is the reduced pressure ($P_R = P/P_C$, where P_C is the critical pressure of the particular substance), and $\tau = 1 - T_R$. The extended Riedel equation is a five-parameter equation:

$$\ln P = A - \frac{B}{T} + C \ln T + DT^2 + \frac{E}{T^2} \qquad (15)$$

where A, B, C, D, and E are adjustable parameters.

The results of this comparison are shown in Table 2. It can be seen that the accuracy of the SINT correlation is significantly higher than that of the three-parameter Antoine equation and only slightly lower than the accuracy of the five-parameter extended Riedel equation. Wagner's equation provides the most accurate correlations for the vapor pressure of oxygen, argon, and nitrogen, but it should be noted that this equation was actually optimized (by stepwise regression) for the very same substances. From the results of this comparison, it can be concluded that the general purpose SINT competes favorably with the most accurate specific vapor-pressure equations for representing data covering a wide temperature range.

Table 3. Data Characterization and Fit of the SINT Equation: Heat Capacity Data

parameter	ammonia (solid) value	ammonia (solid) confidence interval	acetonitrile (solid) value	acetonitrile (solid) confidence interval	butanoic acid (liquid) value	butanoic acid (liquid) confidence interval
a	6.060 926 1	1.72×10^{-4}	7.181 925 1	9.92×10^{-2}	187.575 92	5.79×10^{-2}
b	−0.110 722 8	2.17×10^{-5}	−0.253 187	0.029 361	0.031 031 1	0.015 356 3
c	−2.161 970 5	2.12×10^{-4}	−2.962 873 2	0.186 825	0.346 169 3	0.085 793 5
d	0.323 073 3	2.56×10^{-5}	0.294 771 1	0.012 295	0.052 612 8	0.002 731 7
variance	0.003 805 5		0.019 278 3		0.012 722 8	
sample size (N)	42		30		31	
temp units	K		K		K	
temp range	15.04	191.09	15.63	208.98	272.75[a]	373.06
median T	95.88		62.105		324.4	
heat capacity units	cal/K·mol		cal/K·mol		J/K·mol	
heat capacity range	0.176	11.94	0.409	18.7	169	209
median C_p	6.0015		7.1025		187.2	

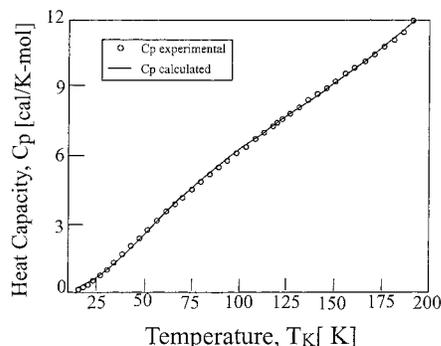[a] An outlying measurement was deleted at $T = 300.18$ K.

Figure 2. Experimental data and calculated curve (using SINT) for solid ammonia's heat capacity.

Heat Capacity of Liquids and Solids. The SINT was fitted to heat capacity data of solid ammonia,[17] solid acetonitrile,[18] and liquid butanoic acid.[19] It should be mentioned that in the case of the butanoic acid one outlying measurement was removed from the data. Solid acetonitrile has a phase transition point at $T = 208.98$ K, and only the data below this temperature were used in the regression. Some additional information for characterization of these data is shown in Table 3. Because heat capacity data are usually modeled using C_p (not the logarithm of C_p) as a dependent variable, eq 13 has been used for modeling the data. The regression results are also shown in Table 3.

It can be seen that the regressed model fits the data well. The variance ranges from 0.0038 for solid ammonia up to 0.019 for solid acetonitrile. All of the confidence intervals on the model parameters are smaller than the parameter values. The experimental heat capacity data and calculated curve are plotted versus the temperature in Figure 2 (for solid ammonia) and that in Figure 3 (for liquid butanoic acid). For the case of butanoic acid, all of the experimental data points are virtually located on the calculated curve, while for the solid ammonia, there are some small discrepancies with the experimental data. It should be emphasized that the shapes of these curves are completely different from the shape of the vapor-pressure curve. For solid ammonia, the curve is moderately concave, and for liquid butanoic acid, the curve is slightly convex.

Heat capacity data for liquids and solids are usually modeled with polynomials (see, for example, ref 9). High-order polynomials are often needed to represent the data accurately, but high-order polynomials can become unstable (indicated by confidence intervals which are larger than the respective parameter values). Even if instability is not detected based on the confidence intervals, often addition of more polynomial terms
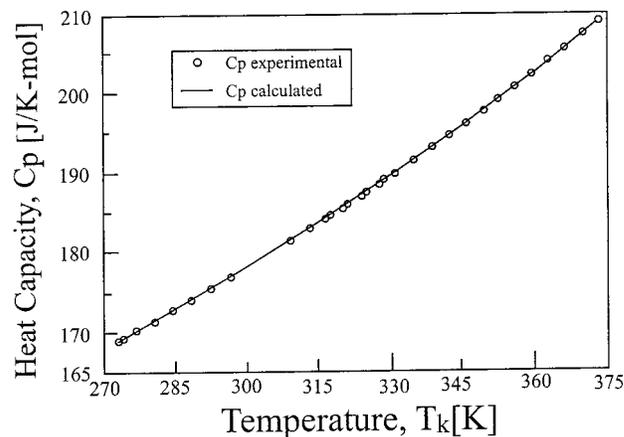

Figure 3. Experimental data and calculated curve (using SINT) for liquid butanoic acid's heat capacity.

Table 4. Comparison of the Heat Capacity Correlation Equations

correlation	variance at optimal solution (normalized to SINT) ammonia	acetonitrile	butanoic acid
fifth-order polynomial	0.0019 (0.4969)	0.0084 (0.4380)	unstable
fourth-order polynomial	0.0026 (0.6772)	unstable	unstable
third-order polynomial	0.0083 (2.1693)	0.0157 (0.8130)	unstable
SINT	0.0038 (1.0000)	0.0193 (1.0000)	0.0127 (1.0000)
second-order polynomial	0.0090 (2.365)	0.1325 (6.8724)	0.0151 (1.1901)

may add unwarranted inflections, resulting in poor interpolation or extrapolation. Use of extreme care is recommended[9] when fitting third- and fourth-order polynomials.

The performance of the SINT for the heat capacity data was compared with the performance of polynomials of various orders. The results of this comparison are summarized in Table 4. In this table the variance and the relative variance (=1 for SINT) of the various fits are shown. The variance of the SINT is smaller than that of the second-order polynomial in all cases. Attempting to fit a higher order polynomial to the data of liquid butanoic acid yields unstable models where the confidence intervals on the parameters are larger than the parameter values. For the case of third-order polynomials, in one case (solid ammonia) the variance is larger than that of the SINT fit and in the other case it is smaller. Increasing the order of the polynomials yields an unstable model in one case and further

Table 5. Shapes of Curves Representing Change of Various Properties as a Function of Temperature

no.	property	change with increasing temp	shape of the curve
1	vapor pressure, liquid	increasing	highly convex
2	heat capacity, solid	increasing	moderately concave
3	heat capacity, liquid	increasing	slightly convex
4	viscosity, gas	increasing	moderately concave
5	viscosity, liquid	decreasing	highly concave
6	surface tension, liquid	decreasing	slightly concave

reduction of the variance in other cases. Thus, in general, SINT yields better fits than low-order polynomials (second or third order). Higher order polynomials may become unstable or enable further reduction of the variance; however, high-order polynomials tend to introduce unwarranted inflections, as we have mentioned earlier.

Fitting SINT to Curves of Various Shapes. To test the capability of the SINT to model the complete variety of curve shapes that appear in modeling of physical and thermodynamic property data (see a detailed demonstration in ref 9), additional properties, namely, the viscosity of vapor and liquid and the surface tension of liquid, were also modeled with the SINT. The quality of the representation was similar to the quality of the representation for vapor pressure and heat capacity data.

The shapes of the various curves that are represented with high precision using the SINT are listed in Table 5. It can be realized that vapor pressure, heat capacity, and vapor viscosity are increasing functions of the temperature, while liquid viscosity and surface tension are decreasing functions. Vapor pressure and liquid heat capacity are convex functions (high convexity for vapor pressure and slight convexity for heat capacity), while viscosity, surface tension, and solid heat capacity are concave functions (high concavity for liquid viscosity, moderate concavity for vapor viscosity and solid heat capacity, and slight concavity for surface tension). Thus, the SINT can accurately model the entire range of function shapes that appear in the representation of physical and thermodynamic data as a function of temperature for pure substances.

7. Conclusions

It has been shown in this paper that the "origin" INT (eq 6) is a generalization of several well-known and widely used physical property correlation equations. The pertinent equations include the Antoine equation for vapor-pressure correlation, the truncated Riedel equation for vapor pressure and liquid viscosity correlation, the Rackett equation for liquid density correlation, and the Guggenheim equation for surface tension correlation. The ability of the INT equation to adjust to so many different equation forms proves its versatility. Furthermore, it shows that the experimental data provide links between physical theory-based models and models based on the statistical distribution of the data. These links can be utilized for a better understanding of the data modeling process.

The SINT equation, one of the derivatives of the "origin" INT, is found to be more convenient for routine use in the regression of physical and thermodynamic data. It has proved to represent properly vapor-pressure data for a temperature range extending from the triple point to the critical point, temperature dependence of solid and liquid heat capacity, vapor and liquid viscosity, and surface tension. Temperature dependencies of these properties show curves of very different shapes such as increasing−decreasing, highly, moderately, or slightly convex, and highly, moderately, or slightly concave. The versatility of the SINT equation to represent curves of different forms makes it an ideal tool to model the temperature dependence of various physical and thermodynamic properties especially in cases where no physical theory-based models exist, when only small amounts of data are available, or when noisy data do not permit fitting of models with many parameters. Because of its limited number of parameters, the new SINT model can be expected to be also accurate and stable in interpolation, extrapolation, and property prediction.

Literature Cited

(1) Reid, R. C.; Prausnitz, J. M.; Poling, R. E. *The Properties of Gases and Liquids*, 4th ed.; McGraw-Hill: New York, 1987.

(2) Wagner, W. New Vapor Pressure Measurements for Argon and Nitrogen and a New Method for Establishing Rational Vapor Pressure Equations. *Cryogenics* **1973**, *13*, 470.

(3) Shacham, M.; Brauner, N. Minimizing the Effects of Collinearity in Polynomial Regression. *Ind. Eng. Chem. Res.* **1997**, *36* (10), 4405−4412.

(4) Shore, H. General Control Charts for Variables. *Int. J. Prod. Res.* **2000**, *38* (8), 1875−1897

(5) Box, G. E. P.; Cox, D. R. An Analysis of Transformations. *J. R. Stat. Soc. B* **1964**, *26*, 211−243.

(6) Belsley, D. A.; Kuh, E.; Welsch, R. E. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*; John Wiley & Sons: New York, 1980.

(7) Shacham, M.; Brauner, N. Identifying and Removing Sources of Imprecision in Polynomial Regression. *J. Math. Comput. Simul.* **1998**, *48* (1), 77−93.

(8) Shacham, M.; Brauner, N. Considering Precision of Experimental Data in Construction of Optimal Regression Models. *Chem. Eng. Process.* **1999**, *38*, 477−486.

(9) Daubert, T. E. Evaluated Equation Forms for Correlating Thermodynamic and Transport Properties with Temperature. *Ind. Eng. Chem. Res.* **1998**, *37*, 3260−3276.

(10) Blau, G. E. *Simusolv Reference Manual*; 1986.

(11) Neter, J.; Wasserman, W.; Kutner, M. H. *Applied Linear Statistical Models. Regression, Analysis of Variance and Experimental Designs*, 3rd ed.; Irwin, Inc.: Burr Ridge, IL, 1990.

(12) Antoine, C. Thermodynamics, Vapor Pressure: A New Relation between Vapor Pressure and Temperature (Thermodynamique, Tensions des Vapeurs: Novelle Relation entre les Tensions et les Temperatures). *C. R. Hebd. Seances Acad. Sci.* **1888**, 116, 117, 681−834, 836, 837, and 1143−1145.

(13) Rackett, H. G. Equation of State for Saturated Liquids. *J. Chem. Eng. Data* **1970**, *15*, 514.

(14) Guggenheim, E. A. The Principle of Corresponding States. *J. Chem. Phys.* **1945**, *13*, 253.

(15) Riedel, L. A New Universal Formula for Vapor Pressure (Eine Neue Universelle Dampfdruckformel). *Chem. Ing. Tech.* **1954**, *26*, 83.

(16) Wagner, W.; Ewers, J.; Pentermann, W. New Vapour-Pressure Measurements and a New Rational Vapour-Pressure Equation. *J. Chem. Thermodyn.* **1976**, *8*, 1049−1060.

(17) Overstreet, R.; Giauque, W. F. Ammonia. The Heat capacity and Vapor Pressure of Solid and Liquid. Heat of vaporization. The Entropy Values from Thermal and Spectroscopic Data. *J. Am. Chem. Soc.* **1937**, *59*, 254−259.

(18) Putnam, W. E.; McEachern, D. M., Jr.; Kilpatrick, J. E. Thermodynamic Properties of Acetonitrile. *J. Chem. Phys.* **1965**, *42* (2), 749−755.

(19) Martin, J. F.; Andon, R. J. L. Thermodynamic Properties of Organic Oxygen Compounds, Part LII. Molar heat capacity of Ethanoic, Propanoic and Butanoic Acids. *J. Chem. Thermodyn.* **1982**, *14*, 679−688.