

# Database Tools for Evaluating Thermophysical Property Data

Richard L. Rowley · W. Vincent Wilding ·  
John L. Oscarson · Yan Yang

Published online: 17 August 2007  
© Springer Science+Business Media, LLC 2007

**Abstract** Most thermophysical-property databases (TPD) provide low-level quality control checks. This manuscript focuses on additional, higher-level data evaluations made possible by the breadth of data stored in the database. For example, thermodynamic equations relate the critical point, vapor-pressure curve, enthalpy of vaporization, liquid density, and liquid and vapor heat capacities to each other. Thermodynamic consistency among these properties can be used to guide selection of the best data sets. Even more broadly, molecular structure-based trends in properties can be identified within the database, and the properties of structurally related compounds can be effectively used to discriminate among available datasets. Automated property predictions can be used in conjunction with the TPD to guide the selection of the most accurate data. These and other high-level consistency tools will be illustrated based on evaluation and quality control work associated with the DIPPR<sup>®</sup> 801 TPD project for pure chemicals.

**Keywords** Database · Quality control · Thermophysical data · Thermophysical properties · Tools

## 1 Thermophysical-Property Data Evaluation and Databases

Thermophysical-property databases (TPDs) are commonly used as educational resources, as property sources for design and problem solving, and as data engines for process design simulators. An often overlooked but important use of a TPD is as a self-improvement tool, a tool to evaluate and enhance the accuracy of the stored data.

---

R. L. Rowley (✉) · W. V. Wilding · J. L. Oscarson · Y. Yang  
BYU-DIPPR Thermophysical Properties Laboratory, Department of Chemical Engineering,  
Brigham Young University, 350 CB, Provo, UT 84602, USA  
e-mail: rowley@byu.edu

The purpose of this article is to suggest ways in which TPDs can be better utilized for this function and to promote continued development of better database quality control tools that utilize the stored data. The tools suggested in this article are from a repertoire of such tools developed under the DIPPR<sup>®</sup> 801 [1] database (Design Institute for Physical Properties, Project 801: Evaluated Process Design Data) project, but they should be broadly applicable to other databases as well.

The objective of the DIPPR<sup>®</sup> 801 database project is to provide *complete* and *evaluated* thermophysical property data for industrially relevant compounds to the practicing engineer and end-user. This is different than more extensive TPDs that seek to provide a comprehensive compilation of the world's thermophysical property data. While the raw data in DIPPR<sup>®</sup> 801 are meant to be exhaustive for the compounds included in the database, only selected industrially important compounds are included in the database; coverage of chemicals is, therefore, modest (approximately 2,000 chemicals) compared to several other databases. This is by design since the first focus of the database staff is providing a recommended value for every constant property (32) and the best temperature correlation for every temperature-dependent property (15) included in the database. Thus, whenever possible, recommended values are given in the database for all 47 properties, and this is what is meant by *complete*. This view of completeness requires that predicted values be used to fill in property values in the absence of experimental data. The second focus of the database staff is the accuracy and reliability of the recommended values as ascertained by all known measures, including the tools described in this article, and this is what is meant by *evaluated*. The evaluation process uses comparisons within chemical families, property interrelationships, and validated and appraised prediction methods in addition to author-reported experimental details (chemical purity, measurement method, experimental procedure, reproducibility, estimated uncertainty, etc.) to determine recommended values. The philosophy is that the database staff is in a better position to determine the most accurate values than the engineer in the field.

## 2 Database Tools for Evaluation

In this article, it is assumed that relational database features that minimize typographical errors, incorrect data type entries, and incorrect signs are commonly used. The focus here is on more advanced property evaluation tools. These tools range from automated checks to very labor-intensive evaluations requiring trained personnel with excellent backgrounds in thermodynamics. Complete evaluation using all of the tools described in this article requires substantial time and effort, but the enhancement in data quality can be significant.

### 2.1 Tool 1: Automated Quality Control Check

The first tool is an automated quality control check (QCC) that includes 39 internal consistency checks. Only data for the specific chemical are used; no cross checking between different chemicals is performed with this first tool. However, the QCC can be

run on a single chemical, groups or families of components, or on the entire database. This facilitates checking properties within groups, such as families, so that insights gained from relationships within the group may be further used to extend the QCC testing and modify the internal consistency levels of the stored property values.

The 39 checks performed by QCC are shown in Table 1. Since these consistency checks compare inter-related properties within the database on a numerical basis, human decision making is minimal. The 39 checks are not all equally important and the underlying inter-relationship may vary in accuracy from exact to only approximate. Standard numerical tolerances are set for each check to provide the desired level of quality control based on the quality of data and the exactness of the inter-relationship. Table 1 classifies the checks into three different “severity” groups: 1 indicates an exact or nearly exact relationship, 2 indicates an inter-relationship known to be effective and reliable for most compounds, and 3 indicates an approximate correlation that has been found useful for screening many errors. When a comparison check falls outside of the set tolerance, the QCC flags the check with a “Failed,” “Warning,” or “Caution” label corresponding to the severity level shown in Table 1. For example, failure of a severity=1 test produces a “Failed” label and the properties involved must be re-evaluated until values are determined that pass this test. On the other hand, failure of a severity=3 test produces the “Caution” label indicating that the value is unusual or out of accepted bounds. Experience with the QCC permits tuning of the tolerances and flag settings so that important discrepancies among properties are not missed and non-substantive cautionary and warning flags do not proliferate to the point of losing significance to the project staff.

## 2.2 Tool 2: Thermodynamic Evaluation

Checks 1, 7, 8, 16, 17, 19, 20, 24, and 31 are useful for evaluating properties in terms of each other and for discriminating among rival data sets. Improved quality is obtained by applying data from additional property measurements, because systematic errors inherent in different property measurement techniques are not likely to be the same. Generally screening of the temperature-dependent data occurs at the dataset level, although later manual evaluation with the tools discussed below can occur for individual data points.

The Clapeyron equation,

$$\frac{d \ln P^*}{d(1/T)} = - \frac{\Delta H_{\text{lat}}}{R \Delta Z_{\text{lat}}} \quad (1)$$

is a primary example of this tool. As expressed in Eq. 1, this exact equation relates the vapor pressure,  $P^*$ , to the latent enthalpy,  $\Delta H_{\text{lat}}$ , and the latent compressibility factor,  $\Delta Z_{\text{lat}}$ . This equation may be applied along the liquid–solid, liquid–vapor, or solid–vapor saturation curves using the appropriate latent enthalpies for fusion, vaporization, or sublimation, respectively. When applied to the liquid–vapor saturation curve, this

**Table 1** Automated quality control checks

#	Severity	Check	Description
1	1	ACEN–VP correlation	Compares stored acentric factor to value calculated from vapor correlation using definition of acentric factor
2	1	$P_t$ –VP correlation	Compares stored triple point pressure with value calculated from vapor pressure correlation at $T_t$
3	1	CAS number verification	Uses check-sum to compare with last digit in CAS number
4	1	Values of properties at $T_{\max}$ and $T_{\min}$	Compares stored values of all $T$ -dependent properties at their lower and upper limits with values calculated from the correlations at $T_{\max}$ and $T_{\min}$
5	1	$S$ , $G$ , $H$ of formation	Checks thermodynamic consistency of formation quantities: Compares $\Delta S_f$ with value calculated from $S_{o, \text{compound}}$ and $S_{o, \text{element}}$ and with $(\Delta H_f - \Delta G_f)/T$
6	2	Flammability limits	(a) Compares stored flammability limits with $T$ 's obtained from vapor pressure equation at partial pressures corresponding to these limits; (b) Calculates the stoichiometric ratio for complete combustion with oxygen, converts that to a ratio with air, and ensures that it is less than the lower flammability limit
7	1	HCOM and HSTD consistency	Compares stored heat of combustion with values calculated from heats of formation for chemical and standard combustion products
8	2	ICP, LCP and HVP consistency	Checks thermodynamic consistency of difference $ICP(T) - LCP(T)$ , calculated from heat capacity correlations at 10 $T$ 's, with $T$ -derivative of HVP correlation
9	1	MW consistency	Compares stored molar mass with value calculated from sum of atomic weights
10	1	LVOL–Density correlation	Compares stored molar volume at $T_{\text{std}}$ with the value obtained from the density correlation
11	1	FP–VP correlation	Compares stored flash-point temperature with value obtained from VP pressure equation at the partial pressure corresponding to lower and upper flammability limits
12	1	Solubility parameter	Compares stored solubility parameter with value calculated from heat of vaporization and density correlations using definition of solubility parameter
13	2	Regression coefficients	Verifies the accuracy of correlation coefficients for all properties by calculating values at the same $T$ at which experimental values are stored in database. Generates AAD, Bias, RMS, and sum of squared residuals for comparison purposes
14	1	Density correlation at $T_c$ and $T_{\text{std}}$	Compares stored values of $V_c$ and molar volume with values calculated from density equation at $T_c$ and $T_{\text{std}}$ , respectively
15	3	VDWA/VDWV	Checks that the ratio of van der Waals area over volume is between $1.0 \times 10^{10} \text{ m}^{-1}$ and $1.7 \times 10^{10} \text{ m}^{-1}$
16	1	VP check at $T_b$	Compares stored normal boiling point with value calculated from vapor pressure correlation at 101,325 Pa
17	1	VP check at $T_c$	Compares recommended $P_c$ value with that obtained from vapor pressure correlation at $T_c$

**Table 1** continued

#	Severity	Check	Description
18	1	$Z_c$ consistency	Compares stored $Z_c$ with value obtained from definition and other critical
19	1	SVP=VP equality at $T_t$	Compares calculated values from the solid and liquid vapor pressure correlations
20	1	VP–HVP Clapeyron check	Compares derivative of VP equation at 10 $T$ 's with values calculated using HVP correlation, LDN correlation, and Soave EOS
21	3	SVR check	Compares magnitudes of stored values with values obtained from Tsonopoulos [10] correlation
22	3	HVP comparison with Watson correlation	Compares exponent computed from HVP correlation in Watson [11] form with 0.38
23	1	HVP=0 at $T_c$ and $T_c$ is $T_{max}$ for VP	Ensures that HVP correlation is zero at $T_c$ and that $T_{max}$ is $T_c$ for VP correlation
24	2	HFOR–HSTD consistency with latent quantities	Compares HSTD–HFOR to HVP( $T_{std}$ ), HFUS( $T_{std}$ ) or HSUB( $T_{std}$ ) for appropriate standard-state phase
25	3	Compare HVP with Chen's correlation	Compares HVP values from correlation with Chen's [12] correlation
26	3	Compare RG with ACEN	Compares RG with Wilding–Rowley [13] correlation of RG with ACEN
27	1	ST=0 at $T_c$	Ensures that ST correlation is zero at $T_c$
28	3	Absolute values VDWa and VDwV	Ensure that the absolute values of VDWa and VDwV are within reasonable magnitudes, $VDwV_{methane} < VDwV < VDwV_{triolein}$ and $VDWa_{methane} < VDWa < VDWa_{triolein}$
29	3	LVS at $T_b$	Ensures that LVS at $T_b$ obeys the DIPPR rule of thumb [14], $0.15 \text{ mPa s} < LVS(T_b) < 0.55 \text{ mPa s}$
30	3	Silva–Hall SVR method	Compares the values of SVR with the predicted value using Silva–Hall [15] method
31	2	Compare HSUB with HVP( $T_t$ ) + HFUS	Compares stored enthalpy of sublimation with values calculated from HVP( $T_t$ ) and HFUS
32	3	RI–DC consistency	Compares dielectric constant to $RI^2$ for nonpolar fluids
33	3	ST–DC consistency [16]	$ST=(20.9 \text{ DC} - 20.5) \text{ dyn/cm}$
34	3	SOLP–DC consistency [16]	$SOLP=(0.22 \text{ DC} + 7.5) (\text{cal/cm}^3)^{1/2}$
35	1	$T_b < T_c$	Boiling point must be below critical temperature
36	2	$T_m < T_b$	Melting point generally below boiling point
37	1	Liquid range check	Check that $T_{min}$ is $\geq T_t$ and $T_{max} \leq T_c$
38	3	$Z_c$ range check	Check that $0.1 < Z_c < 0.6$
39	3	ACEN range check	Check that the acentric factor is between 0 and 2

AAD, average absolute deviation; ACEN, acentric factor; CAS, chemical abstracts service; FP, flash point; HCOM, enthalpy of combustion; HFOR, ideal gas enthalpy of formation; HFUS, enthalpy of fusion; HSTD, standard state enthalpy of formation; HSUB, enthalpy of sublimation; HVP, heat of vaporization; ICP, ideal gas heat capacity; LCP, liquid heat capacity; LVOL, liquid volume at  $T_{std}$ ; LVS, liquid viscosity; MW, molar mass;  $P_t$ , triple point pressure; RG, radius of gyration; RMS, root mean square; ST, surface tension; SVR, second virial coefficient;  $T_b$ , normal boiling point;  $T_c$ , critical temperature;  $T_{max}$ , maximum valid temperature for correlation;  $T_{min}$ , minimum valid temperature for correlation;  $T_{std}$ , 298.15 K;  $T_t$ , triple point temperature; VDWa, van der Waals area; VDwV, van der Waals volume; VP, vapor pressure

equation explicitly ties together the temperature dependence of three properties<sup>1</sup> in the DIPPR<sup>®</sup> database: the heat of vaporization  $\Delta H_{\text{vap}}$ , the vapor pressure, and the liquid molar density,  $\rho$ . The liquid density is of course related to the liquid compressibility factor by

$$Z = \frac{P}{\rho RT}. \quad (2)$$

Equation 1 also implicitly ties the ideal-gas heat capacity,  $C_{P,\text{IG}}$ , and the liquid heat capacity at constant pressure,  $C_{P,\text{L}}$  to the vapor pressure through

$$\frac{d\Delta H_{\text{vap}}}{dT} = C_{P,\text{IG}} - C_{P,\text{L}} + \Delta C_{P,\text{corr}}^{\text{dep}} + \Delta C_{P,\text{corr}}^{\text{sat}} \quad (3)$$

The third term on the right side of this equality is a heat-capacity departure function that corrects for gas-phase nonidealities. It can be evaluated with an equation of state using

$$\Delta C_{P,\text{corr}}^{\text{dep}} = T \int_{\infty}^V \left( \frac{\partial^2 P}{\partial T^2} \right)_V dV - T \left( \frac{\partial P}{\partial T} \right)_V \left( \frac{\partial P}{\partial V} \right)_T^{-1} - R \quad (4)$$

The final term in Eq. 4 converts between a heat capacity along the saturation curve to a constant-pressure heat capacity. It can be evaluated using

$$\Delta C_{P,\text{corr}}^{\text{sat}} = \left[ V_{\sigma}^{\text{vap}} - T \left( \frac{\partial V^{\text{vap}}}{\partial T} \right)_P \right] \left( \frac{dP}{dT} \right)_{\sigma} \quad (5)$$

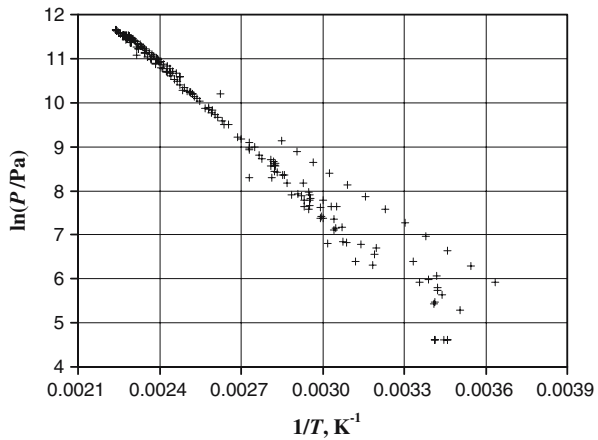
These two correction terms are inconsequential below a reduced temperature of about 0.65, approximately the normal boiling point.

As a first case study for the thermodynamic consistency tool, consider the task of recommending the best correlation for the vapor pressure (VP) and the best value for the normal boiling point ( $T_b$ ) for tetraethoxy silane. The vapor-pressure data available in the DIPPR<sup>®</sup> 801 database for this compound are shown in Fig. 1. There is considerable disagreement among the 27 datasets, particularly at low temperatures. The critical temperature and pressure have been accurately measured and may be accepted as reliable. There is an 8 K range in the 39 reported normal boiling points for this compound. The different slopes of the VP datasets suggest systematic errors in some of the measurements. Consider the following four different choices for the recommended vapor-pressure correlation and how  $\Delta H_{\text{vap}}$  and  $C_{p,\text{L}}$  data help guide the choice.

In choice (a), we assume that the high-temperature dataset with the lower slope and the matching dataset at lower temperatures are more accurate than the others. The VP correlation regressed using only those data points are shown in Fig. 2a. The resultant curves for  $\Delta H_{\text{vap}}$  and  $C_{p,\text{L}}$  obtained from Eqs. 1 and 3 can be compared to available

<sup>1</sup> It also ties the vapor density to these properties, but the vapor density is a function of pressure as well as temperature and is not included in the DIPPR database. Generally, the DIPPR<sup>®</sup> 801 staff uses the Soave equation of state when applying Eq. 1 for the vapor volume.

**Fig. 1** Vapor pressure data available in the DIPPR<sup>®</sup> 801 database for tetraethoxysilane



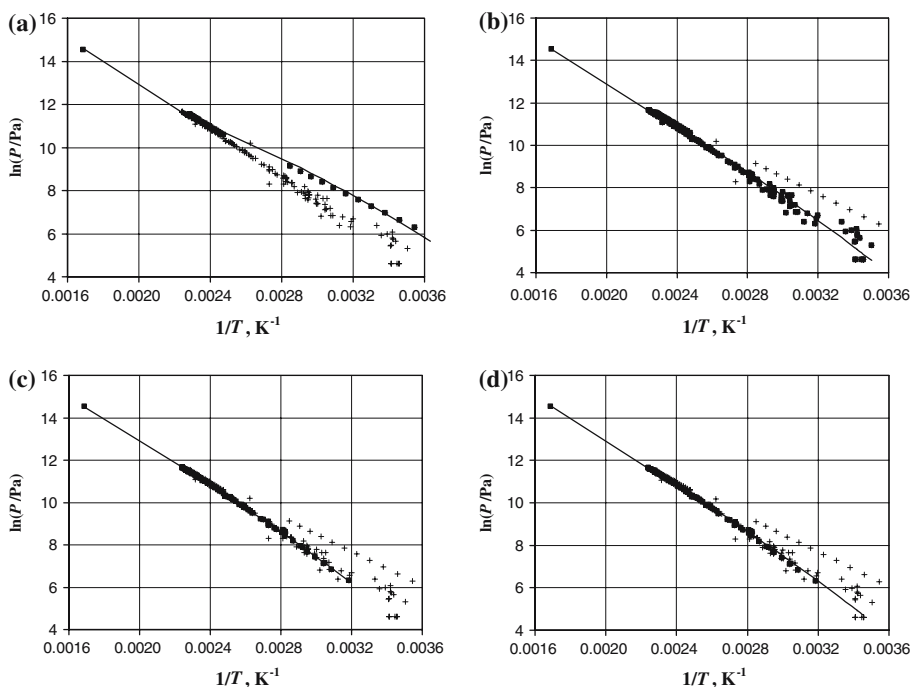
experimental data. Figure 3a shows that, despite large scatter in the reported data, the slope of the derived  $\Delta H_{\text{vap}}$  curve is too steep, suggesting that this was not the best choice. Figure 4a also confirms that this was a poor choice of VP data, because the resultant  $C_{p,L}$  curve is clearly inconsistent with the experimental data.

In choice (b), we assume that the somewhat outlying, higher-pressure data in the low-temperature range are in error. Excluding these data gives a fit of the remaining VP data as shown in Fig. 2b. The effects of this choice on  $\Delta H_{\text{vap}}$  and  $C_{p,L}$  are shown in Figs. 3b and 4b, respectively. Though the scatter in reported  $\Delta H_{\text{vap}}$  data precludes a definitive decision from Fig. 3b, the agreement with  $C_{p,L}$  is clearly substantially improved by this choice of VP data.

In the third choice, we selectively eliminated additional datasets based on reported purity, precision of measurements, and consistency trends. A regression of these data is shown in Fig. 2c. The  $\Delta H_{\text{vap}}$  curve obtained from the regressed VP correlation, shown in Fig. 3c, is in good agreement with the reported  $\Delta H_{\text{vap}}$  data. The corresponding  $C_{p,L}$  curve in Fig. 4c is, however, shifted from the experimental values.

The fourth choice uses the same selected datasets as in the third choice. However, in this case, we perform a simultaneous regression on the VP and  $C_{p,L}$  datasets. When appropriately weighted, the  $C_{p,L}$  curve (Fig. 4d) can be brought into better agreement (in this case the maximum error is  $-3.7\%$  which occurs near the melting point) with the experimentally observed values without much degradation in the VP fit (Fig. 2d), while still providing agreement with the  $\Delta H_{\text{vap}}$  data (Fig. 3d). The resultant VP correlation gives a value of 441.34 K for  $T_b$ .

Figure 5 shows that the best choice for  $T_b$  is fairly sensitive to the literature values of  $C_{p,L}$ . The curves shown are for normal boiling point values of  $T_b = 438, 440.9, 441.34,$  and  $442$  K. Although the curve obtained for  $T_b = 441.34$  K (obtained from the simultaneous regression in choice four) matched the low-temperature experimental  $C_{p,L}$  values to within  $3.7\%$ , the agreement is the worst at the highest temperatures where the correction terms (Eqs. 4, 5) are very small. However, the  $C_{p,L}$  curve generated with  $T_b$  set at 440.9 K agrees with the higher-temperature experimental data, appears to extrapolate to higher temperatures consistent with the experimental trend



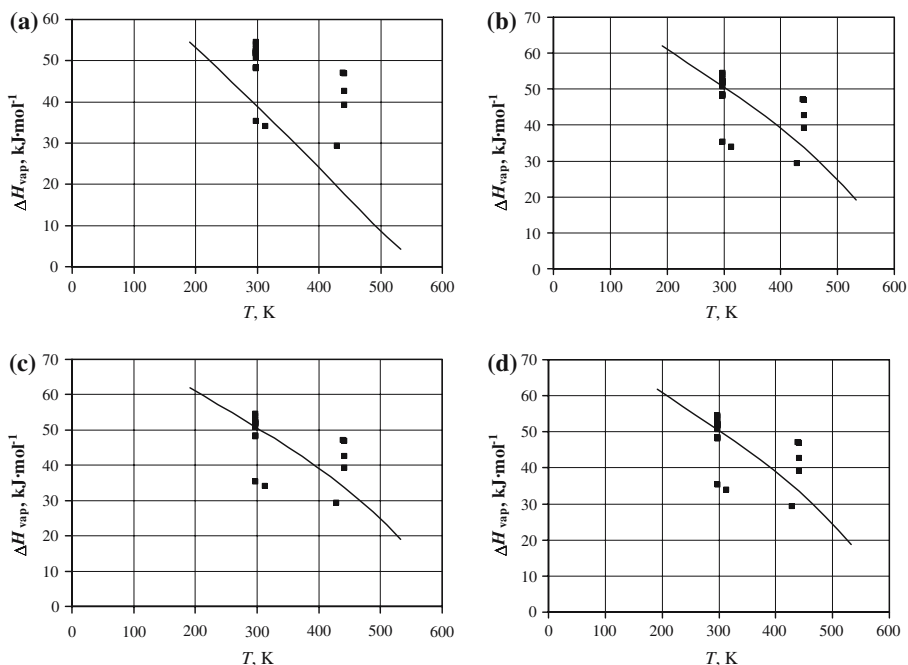
**Fig. 2** Four choices for regression of tetraethoxy silane vapor-pressure data: (a) accepting (■) only the higher vapor-pressure data sets, (b) neglecting (+) the highest of the low-pressure data sets, (c) selecting experimental data sets based on experimental evaluations and internal consistency, and (d) using the selected data set and simultaneously regressing VP and  $C_{p,L}$  data

in the data and known behavior, and only increases the error at the lowest temperature to  $-6.5\%$ . On the other hand, setting  $T_b = 438.0$  K produces a bias of  $-11\%$ . There is not an exact match between the observed temperature dependence of  $C_{p,L}$ , suggesting the need for an improved VP equation with more flexibility in the low-temperature range, but even with this limitation, the use of multiple properties in the evaluation has led to a refined selection of VP data and the best  $T_b$  value.

As a second case study on the use of the thermodynamic consistency tool, consider evaluation of the properties for bicyclo[2.2.1]heptane. Available data include an accurate value for the melting point temperature, six normal boiling point values ranging from 373.15 to 383.15 K, seven enthalpies of sublimation ( $\Delta H_{\text{sub}}$ ), and very accurate solid vapor-pressure data as shown in Fig. 6. The thermodynamic consistency tool can be used with these data to obtain recommended values for  $T_b$ ,  $\Delta H_{\text{sub}}$ , and  $\Delta H_{\text{fus}}$  (the enthalpy of fusion), as well as recommended correlations for the liquid and solid vapor-pressure curves.

Low pressures along the solid–vapor equilibrium curve allow one to set  $\Delta Z = 1$  in Eq. 1 without loss of accuracy. Since  $\Delta H_{\text{sub}}$  is nearly constant over a substantial temperature range, a linear relationship between  $\ln(P)$  and  $1/T$  can be regressed from the solid vapor-pressure data as shown by the solid line in Fig. 6. In doing so, the three lowest-temperature points were neglected, because the corresponding pressures were





**Fig. 3** Curves for  $\Delta H_{\text{vap}}$  obtained from the four VP regressions given in Fig. 2

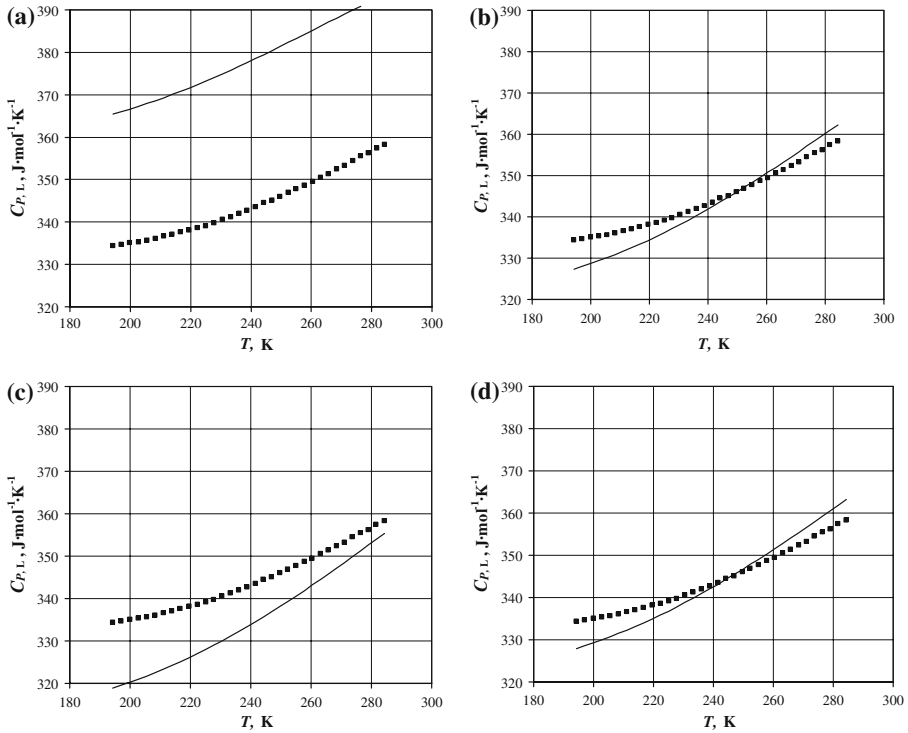
nominal values known only to one significant figure. The slope of the resultant solid vapor-pressure correlation gives  $\Delta H_{\text{sub}} = 40.52 \text{ kJ}\cdot\text{mol}^{-1}$ .

Liquid and solid vapor pressures must meet at the triple point. The regressed solid vapor-pressure curve provides the triple-point pressure at the known melting point (which is assumed to be equal to the triple-point temperature). The Riedel vapor-pressure equation can be used in a predictive mode with known values for  $T_c$  and one other saturation temperature. Requiring the vapor-pressure curve to go through the triple point, as shown in Fig. 6, yields  $T_b = 378.15 \text{ K}$ .

Thermodynamic consistency also requires that

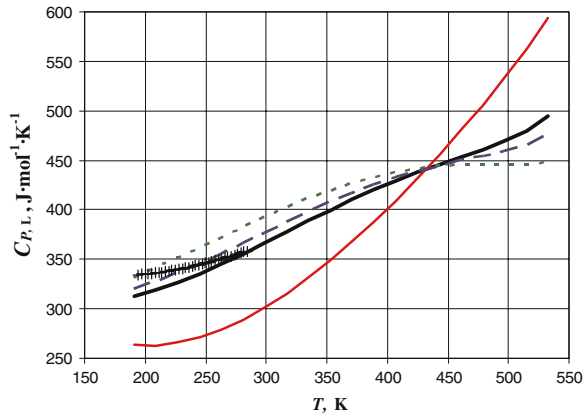
$$\Delta H_{\text{fus}} = \Delta H_{\text{sub}} - \Delta H_{\text{vap}} \quad (6)$$

at the triple point. One can obtain  $\Delta H_{\text{vap}}$  from the vapor-pressure curve using Eq. 1. This gives  $\Delta H_{\text{fus}} = 8.34 \text{ kJ}\cdot\text{mol}^{-1}$ . Although we do not have experimental data for this property, this value is in good agreement with the experimental value for the structurally similar compound bicyclo[2.2.2]octane, for which the measured  $\Delta H_{\text{fus}} = 8.35 \text{ kJ}\cdot\text{mol}^{-1}$ . This comparison of values for structurally similar compounds is a form of the next tool, Tool 3.



**Fig. 4** Curves for  $C_{p,L}$  obtained from the four VP regressions given in Fig. 2

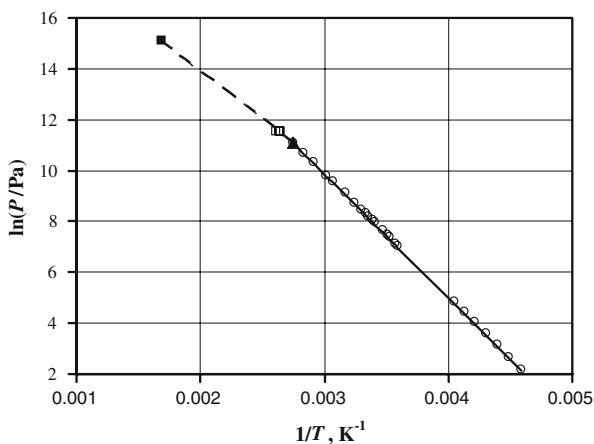
**Fig. 5** Impact of fixed  $T_b$  in the vapor-pressure equation on  $C_{p,L}$ :  $T_b = 438$  K (—),  $T_b = 440.9$  K (—, heavy),  $T_b = 441.34$  K (---),  $T_b = 442$  K (----), experimental points (+)



### 2.3 Tool 3: Molecular Structure-Based Evaluation

Just as multiple inter-related properties can provide additional consistency information with which to hone the selection of the recommended values, groups of inter-related compounds can provide additional information to discriminate among conflicting data

**Fig. 6** Available bicycle[2.2.1]heptane data for solid (o) vapor pressure, the normal boiling point ( $\square$ ), the triple point ( $\blacktriangle$ ), and the critical point ( $\blacksquare$ ) which are used to obtain the solid (—) and liquid (---) vapor-pressure curves



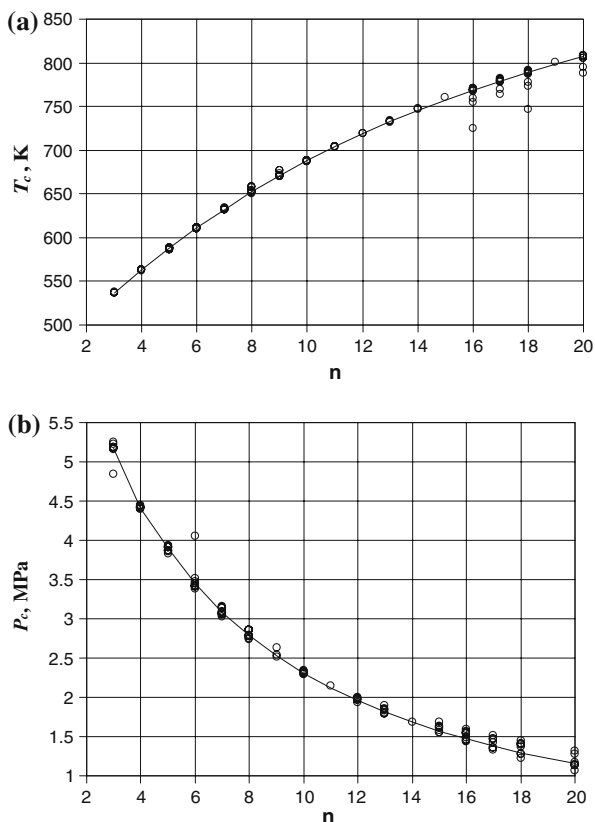
or even enhance experimentally determined values. The underpinning logic of this tool is twofold: structurally similar compounds exhibit similar properties and systematic increments to the molecular structure produce systematic increments to the properties.

As a case study for this tool, consider assigning accurate values for the critical properties of the higher molar mass  $n$ -alcohols. Accurate measurements of the critical properties for these alcohols are difficult, because chemical decomposition occurs below  $T_c$  increasingly with higher molar mass. Nikitin et al. [2] reported rapid, transient measurements of  $T_c$  and critical pressure ( $P_c$ ) for  $C_{13}$  to  $C_{22}$  in 1998, and those measurements have helped establish  $T_c$  for this family of compounds. The available experimental values for  $T_c$  are shown in Fig. 7a as a function of carbon number,  $n$ . The  $T_c$  data were analyzed in conjunction with the experimental uncertainties to obtain internal consistency within the homologous series and, thereby, recommend the best smoothed values shown as a solid line in the figure. Figure 7a shows that improved accuracy can be expected for the recommended values particularly for  $C_{15}$  to  $C_{20}$  due to the guidance provided by additional measurements on compounds within the family. Similarly, available  $P_c$  values are shown in Fig. 7b along with the line representing the smoothed values recommended based on the family trend. The plot suggests an improvement in accuracy of the recommended values for  $C_2$  to  $C_9$  and for  $C_{15}$  to  $C_{20}$ . Critical constants for  $C_{19}$  can be assigned with relative confidence even though they have not been measured.

Critical volumes are difficult to measure and few experimental data are available for the longer-chain alcohols. Figure 8a shows the available data.<sup>2</sup> From the data it is difficult to distinguish between two possible trends for  $V_c$ . However, a plot of critical density versus carbon number shows that the trend displayed as a solid line gives a critical density that in the limit of large carbon number approaches the known behavior of  $n$ -alkanes. This is not true for the lower trend in  $V_c$ . The critical density of  $n$ -alcohols is expected to behave like their  $n$ -alkane backbone in the limit of long

<sup>2</sup> The data represented with diamonds are listed as experimental in the DIPPR database as obtained from another data source. However, they are likely predicted values.

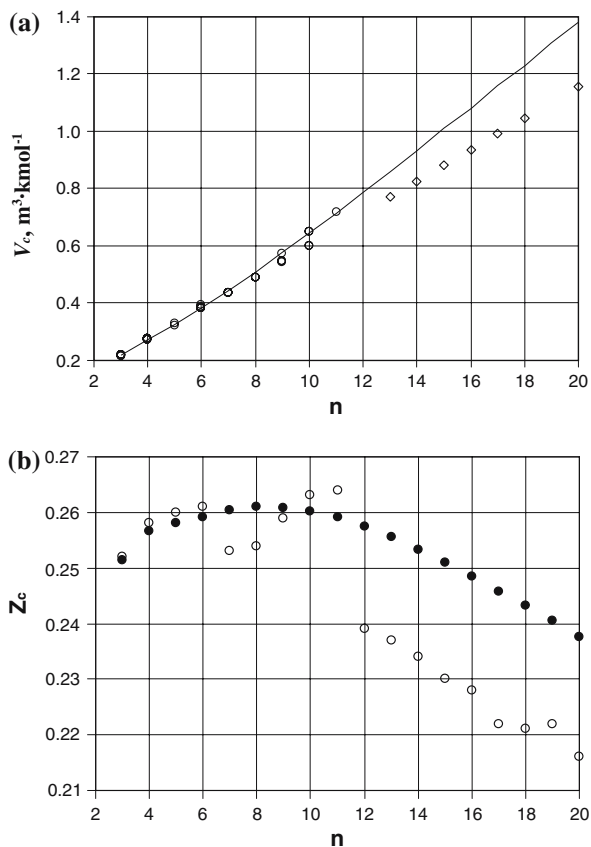
**Fig. 7** Critical temperature (a) and pressure (b) as a function of carbon number ( $n$ ) for  $n$ -alcohols: experimental data (○) and smoothed values (—) recommended in the DIPPR database



chain length. Using the molecular structure tool also improves the correlation of  $Z_c$  (as calculated from the accepted values  $T_c$ ,  $P_c$ , and  $V_c$ ) as shown in Fig. 8b. This figure compares the original recommendations for  $Z_c$  to those made after applying the molecular structure tool. The more logical smooth trend in  $Z_c$  results from refinement of the systematic trends in the individual  $T_c$ ,  $P_c$ , and  $V_c$  values that comprise it and no additional smoothing was performed. Similar improvements can be shown for other properties.

The previous case study illustrates the use of the molecular structure tool when data for a homologous series are available, but the tool can also be used if experimental data are available for a single, similar, reference compound. For example, measurements on  $n$ -hexanal give  $T_c = 591.0$  K. The molecular structure tool can be used to accurately estimate  $T_c$  for  $n$ -heptanal utilizing  $n$ -hexanal as a reference compound.  $n$ -Heptanal differs in structure from the reference compound only by a  $-\text{CH}_2-$  in an aliphatic chain. The difference between the  $T_c$  values predicted by the Lydersen [3] method for these two compounds is 24.92 K. If this increment is added to the experimental value for  $n$ -hexanal, one obtains  $T_c = 615.9$  K for  $n$ -heptanal. In fact, there is an experimental value of 616.8 K  $n$ -heptanal. The more similar the compound is to the

**Fig. 8** Critical volumetric properties for  $n$ -alcohols as a function of carbon number: (a) critical volume experimental data ( $\circ$ ), reported data from a database ( $\diamond$ ), and recommended values based on family trends (-); (b) critical compressibility factor previously recommended in the DIPPR<sup>®</sup> 801 database ( $\circ$ ) and those recommended based on family trends in  $T_c$ ,  $P_c$  and  $V_c$  ( $\bullet$ ).



chosen reference, the better accuracy that can be expected by using this tool in this manner.

#### 2.4 Tool 4: Property Prediction

An automated property prediction package attached to a TPD is a powerful tool. One such package is DIADEM [4] (DIPPR<sup>®</sup> Information And Data Evaluation Manager) which was developed specifically for the DIPPR<sup>®</sup> 801 database but can be used also as a stand-alone application. DIADEM can parse SMILES formulas into structural groups compatible with most group contribution methods. While the prediction package obviously helps fill the completeness requirement of the DIPPR<sup>®</sup> 801 database by providing predicted values in the absence of experimental data, it can also enhance the quality of the recommended values. For example, by comparing predicted and recommended values for all compounds in the database, DIADEM can identify which prediction method works best overall and which methods yield the best results for different families or classes of compounds. For example, Table 2 shows an abbreviated version of the results obtained from DIADEM's evaluation of three methods for

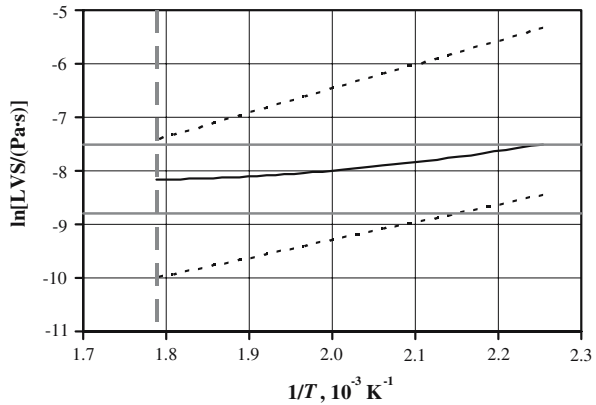
**Table 2** Subsection of table produced by DIADEM comparing three prediction methods for  $T_c$  (average absolute percent deviation, AAD, of predicted values is calculated relative to the recommended values in DIPPR<sup>®</sup> 801)

Family	No. of points	AAD (%)		
		Joback [5]	Lydersen [3]	Wilson–Jasperson [6]
Overall	462	1.42	1.54	2.05
1-Alkenes	15	0.21	0.38	0.16
2,3,4-Alkenes	8	0.63	0.22	0.17
Acetates	10	0.47	0.79	0.39
Aldehydes	9	1.77	2.05	1.29
Aliphatic ethers	16	0.52	0.48	0.74
Alkylcyclohexanes	6	1.16	1.28	2.33
Aromatic alcohols	16	1.30	1.30	0.99
Aromatic amines	13	0.97	1.06	0.94
Aromatic chlorides	3	0.55	0.58	0.45
C, H, F compounds	32	1.17	1.53	1.66
C, H, multihalogen compounds	21	0.63	0.56	0.82
C <sub>1</sub> /C <sub>2</sub> aliphatic chlorides	10	0.85	0.83	0.56
Dimethylalkanes	11	0.43	0.39	1.37
Epoxides	6	1.02	0.88	2.10
Formates	5	1.52	0.71	1.94
Ketones	26	1.51	1.67	8.95
Mercaptans	5	0.37	0.46	0.17
Methylalkanes	11	0.35	0.41	0.53
<i>n</i> -Alcohols	15	2.48	2.66	0.97
<i>n</i> -Aliphatic acids	8	1.30	1.02	1.26
Other aliphatic alcohols	22	2.17	2.27	0.93
Other alkanes	18	0.53	0.60	2.66
Propionates and butyrates	11	0.51	0.57	0.45
Silanes/siloxanes	23	4.54	5.30	5.26
Sulfides	10	1.59	1.25	1.56

predicting  $T_c$ . While the overall average absolute deviation suggests that the Joback [5] method should generally be preferred over that of Lydersen [3] and Wilson–Jasperson [6], the individual family results allow selection of a method that has been better tuned to specific families of compounds. In this case, the Wilson–Jasperson method might be chosen for alkenes, acetates, aldehydes, alcohols, and chlorides; the Lydersen method for sulfides, *n*-aliphatic acids, formates, and epoxides; and the Joback method for other compounds. Matching structure with the technique’s reliability will enhance the quality of the recommended value.

Correlations and prediction methods can also be used to guide the selection of recommended property values, whether that selection is made from disparate exper-

**Fig. 9** Liquid viscosity “rule of thumb” (horizontal gray lines) for the viscosity at  $T_b$  (vertical gray dashed line) guides the selection of the recommended (solid line) correlation from alternative (dashed lines) correlations



imental values or from several predictions of comparable accuracy. Consider for example, the choice between the three predicted liquid viscosity curves shown in Fig. 9. An empirical observation or “rule of thumb” developed for the DIPPR<sup>®</sup> 801 database is that the liquid viscosity of most substances is between 0.15 and 0.55 mPa·s at the normal boiling point. Although there is no current theoretical reason for this correlation, it works for 95% of the accepted experimental data in the DIPPR<sup>®</sup> 801 database. Applying this “rule of thumb” to the predictions shown in Fig. 9 suggests that the correlation shown with the solid line should be recommended over the two represented with dotted lines. This tool has been used to select between competing experimental datasets and has guided the selection of the best viscosity correlation for over 250 compounds in the DIPPR<sup>®</sup> 801 database.

As a second example of how correlations can guide the selection of recommended database values, consider the case of 2,3-butanediol for which  $T_c$  is known but  $V_c$  has not been measured. However, liquid densities have been measured for this compound over a 60 K temperature range. These data coupled with the observation that the Rackett [7] equation generally correlates liquid density very well from the melting point to the critical point allows one to use the Rackett equation to guide the extrapolation of the liquid-density data to the critical point. When written in the form,

$$\rho = \frac{P_c R / T_c}{Z_c^{[1+(1-T_r)]^{2/7}}}, \quad (7)$$

the Rackett equation can be used to fit the density data by adjusting  $Z_c$ , from which  $V_c$  is then calculated. This guided extrapolation gives a value of  $V_c = 0.267 \text{ m}^3 \cdot \text{kmol}^{-1}$ . We believe this value to be more reliable than the either of the individual values of 0.290 or 0.292  $\text{m}^3 \cdot \text{kmol}^{-1}$  predicted by the Ambrose [8] and Fedors [9] methods, respectively, because our testing for compounds where experimental data for  $V_c$  are available has shown that this use of Eq. 7 to extrapolate volumetric data for the given compound is quite reliable.

Table 3 Examples of data triage using impact factors

Compound	$T_c$		$P_c$		$V_c$		$T_b$	
	$\Delta T_c$ (K)	Percent	$\Delta P_c$ (kPa)	Percent	$\Delta V_c$ ( $\text{m}^3 \cdot \text{kmol}^{-1}$ )	Percent	$\Delta T_b$ (K)	Percent
<i>cis</i> -2-Heptene	{4.80}	0.88						
Isopropyl acetate	-1.84	-0.35	229	6.96	{0.0090}	2.69	0.25	0.07
<i>n</i> -Pentyl acetate	-2.10	-0.35	-85	-3.07	{0.0280}	6.33	0.75	0.18
Benzaldehyde	-70.00	-10.07	{-2.470}	-53.12				
Nonanal	{1.70}	0.00	{-50}	-1.83	{0.0160}	3.04		
Methyl <i>tert</i> -pentyl ether	3.00	0.56	{151}	4.97	-0.0090	-2.33	-0.18	-0.05
Ethylcyclohexane	{-5.15}	-0.85	{210}	6.91			0.03	0.01
2,6-Dimethylpyridine	0.25	0.04	{100}	2.67	{-0.0122}	-3.27	-0.29	-0.07
Dimethyl isophthalate	{2.00}	0.26	{-140}	-5.02	{0.0288}	5.34		
1,1,1,2,3,3,3-Heptafluoropropane	1.12	0.30	{76}	2.61	0.0192	6.99	0.86	0.33
Bromotrichloromethane	{-0.60}	-0.10	{-390}	-7.85	{0.0130}	4.58		
Dibromodifluoromethane	{-6.70}	-1.40	{460}	11.30	{-0.0020}	-0.80		
1,2-Dichloropropane	{6.00}	1.05	{410}	9.67	{-0.0013}	-0.44		
Pyrrole	-0.10	-0.02	{-200}	-3.06	{-0.0300}	-13.04		
Ethyl acetate	25.50	4.87	440	11.34	0.0084	2.94	-0.05	-0.01
Isopentyl acetate	13.23	2.26					0.15	0.04
Methyl acetate	45.25	8.93	1,090	22.95	-0.0052	-2.28	-0.24	-0.07
<i>n</i> -Propyl acetate	-12.13	-2.21	170	5.06	0.0072	2.09	-0.44	-0.12
Acetaldehyde	-11.40	-2.45						



**Table 3** continued

Compound	$T_c$		$P_c$		$V_c$		$T_b$	
	$\Delta T_c$ (K)	Percent	$\Delta P_c$ (kPa)	Percent	$\Delta V_c$ ( $\text{m}^3 \cdot \text{kmol}^{-1}$ )	Percent	$\Delta T_b$ (K)	Percent
Methyl ethyl ether	3.80	0.87	288	6.55	-0.0253	-11.43	-0.06	-0.02
Methylcyclohexane	-3.10	-0.54	-10	-0.29			0.04	0.01
Phenol	-0.05	0.00			0.0000	0.00	-0.07	-0.02
<i>m</i> -Cresol	-0.85	-0.12			-0.0020	-0.64	0.04	0.01
<i>o</i> -Cresol	-2.55	-0.37	-150.000	-2.99	0.0000	0.00		

Entries in {} have the potential to replace predicted values with new experimental measurements; otherwise, the entry shows the deviation of the new value from the currently accepted value. All entries show the potential change relative to the currently accepted value

## 2.5 Tool 5: New Data Triage

Keeping an evaluated database up-to-date is a major challenge because of the rate at which new data are published and the interconnection between properties that has been described in Tools 1–3. This enforced consistency between inter-related properties means that updates in key properties may necessitate re-evaluation and upgrade of inter-related properties. This improves the quality of not only the new property value, but also those related properties that are upgraded as a result of the change. The process is nevertheless time consuming, and tools within the database can be useful for identifying what changes need to be made and their effect on other properties. A key component in the process is identification of the “impact factor” (IF) that new data will have on the database. The IF provides a quick triage of the importance of updating the database with the new data.

The IF is determined by comparing unevaluated data put into the database to the property values and correlations currently recommended. Weighting factors in calculation of the IF include the potential to replace predicted values with experimental, the potential to improve the accuracy of accepted values, the magnitude of the difference between the new and accepted values, and the relative quantity of data being added. For temperature-dependent properties, an average absolute deviation of the new values from the current DIPPR<sup>®</sup> 801 correlation is used to help establish the IF. A few illustrative examples have been compiled in Table 3. The { } notation shows that a measured value for  $T_c$  of *cis*-2-heptene is now available to potentially replace the previously predicted value. Similarly, nonanal has a large IF because measured values are given high priority even when they agree well with previously predicted values (in this case differences of only 0.0%, –1.8%, and 3.04% for  $T_c$ ,  $P_c$ , and  $V_c$ , respectively). Ethyl acetate also has a high IF since the new critical measurements differ significantly from the recommended values based on previous measurements (differences of 4.9%, 11.3%, and 2.9% for  $T_c$ ,  $P_c$ , and  $V_c$ , respectively). On the other hand, the IF for compounds like phenol and *m*-cresol would be very low because new measurements of  $T_c$  and  $P_c$  are nearly identical to values already in the database.

## 3 Conclusion

The TPD itself provides powerful tools for evaluating data, improving accuracy, and enhancing the quality of the data. The five tools presented in this article illustrate ideas for doing this, but they are not meant to be exhaustive. The purpose of this article is to engender other ideas that utilize the wealth of information stored in a TPD for self improvement. The five tools represent some of the more important tools within a larger collection that have been developed by and used in the DIPPR<sup>®</sup> 801 database project.

**Acknowledgment** Support of this work under the DIPPR<sup>®</sup> 801 project (*Data Compilation of Pure Chemical Properties*) is gratefully acknowledged.

## References

1. R.L. Rowley, W.V. Wilding, J.L. Oscarson, Y. Yang, N.A. Zundel, T.E. Daubert, R.P. Danner, *DIPPR® Data Compilation of Pure Chemical Properties* (Design Institute for Physical Properties, AIChE, New York, 2006)
2. C. Tsonopoulos, *AIChE J.* **24**, 1112 (1978); *AIChE J.* **20**, 263 (1974); *AIChE J.* **21**, 827 (1975)
3. K.M. Watson, *Ind. Eng. Chem.* **23**, 360 (1931)
4. N.H. Chen, *J. Chem. Eng. Data* **10**, 207 (1965)
5. W.V. Wilding, R.L. Rowley, *Int. J. Thermophys.* **7**, 525 (1986)
6. *DIPPR Policies and Procedures Manual*, BYU-TPL, Brigham Young University, Provo, Utah (2006)
7. G.A. Iglesias-Silva, K.R. Hall, *Ind. Eng. Chem. Res.* **40**, 1968 (2001)
8. A.L. Horvath, *Physical Properties of Inorganic Compounds* (Crane, Russak, New York, 1975)
9. E.D. Nikitin, P.A. Pavlov, A.P. Popov, *Fluid Phase Equilib.* **149**, 223 (1998)
10. A.L. Lydersen, *Estimation of Critical Properties of Organic Compounds, Report 3* (University of Wisconsin, College of Engineering, Experimental Station, Madison, Wisconsin, 1955)
11. J.R. Rowley, R.J. Rowley, W.V. Wilding, J.L. Oscarson, R.L. Rowley, *DIADDEM 2004: DIPPR Information and Data Evaluation Manager, BYU-DIPPR TPL* (Brigham Young University, Provo, Utah, 2006)
12. K.G. Joback, M.S. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts (1984)
13. G.M. Wilson, L.V. Jaspersen, *AIChE Spring Mtg* (New Orleans, Louisiana, 1966)
14. C.F. Spencer, R.P. Danner, *J. Chem. Eng. Data* **17**, 236 (1972)
15. D.L. Ambrose, *Natl. Phys. Lab. Report Chem.* **98**, (1979)
16. R.F. Fedors, *AIChE J.* **25**, 202 (1979)