International Journal of Thermophysics, Vol. 25, No. 2, March 2004 (© 2004)

State Behavior Database for Pure Liquids and Data Correlation[1]

I. Cibulka[2,3] and T. Takagi[4]

A database system designed for storage, maintenance, and evaluation of published experimental data related to the state behavior of pure liquids is described. The database system is maintained by a program package (FoxPro) that allows the effective and user-friendly process to enter, store, handle, and update the information starting from a bibliography and ending in a data evaluation. Evaluation of data is performed using either two functions of temperature for data at 0.1 MPa or at saturation or the Tait equation for density data at elevated pressures. The weighted least-squares method is employed to evaluate the adjustable parameters of smoothing functions. Search and export program modules allow various views, searches, and outputs of the content of the databases. At present, the database system contains about 4,400 records in the bibliographic part, more than 100,000 experimental data points for about 500 substances, and over 800 sets of parameters of smoothing functions for about 330 substances. The database system is intended as an effective tool for compilation and critical evaluation of data, not as a commercial databank for distribution.

KEY WORDS: correlation; database; density; liquids; state behavior.

1. INTRODUCTION

Data on the state behavior of fluids are, besides data on thermal properties (heat capacities), some of the most important information necessary to describe the thermodynamic properties of fluid systems. Tremendous

---

[1] Paper presented at the Fifteenth Symposium on Thermophysical Properties, June 22–27, 2003, Boulder, Colorado, U.S.A.

[2] Department of Physical Chemistry, Institute of Chemical Technology, Technická 5, 166 28 Prague, Czech Republic.

[3] To whom correspondence should be addressed. E-mail: ivan.cibulka@vscht.cz

[4] Kyoto Institute of Technology, Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan.

progress in computer technologies over the last several years has enabled the development of extensive databases and databanks for effective storage, maintenance, and evaluation of large amount of numerical and textual data. In the course of the past ten years a complex database was developed with the cooperation of laboratories in Prague and Kyoto that stores, handles, and outputs data related to the state behavior of pure organic liquids. Values for more than 100,000 data points are collected. A significant portion of the density data have been critically evaluated into a form of empirical functions of temperature or temperature and pressure, and the results have been published: densities at atmospheric pressure or at saturation of 1-alkanols from C_1 to C_{10} and n-alkanes from C_5 to C_{16} [1], compressed-liquid densities of 1-alkanols from C_1 to C_{10} [2], n-alkanes from C_5 to C_{16} [3], ethers, carbonyl compounds, organic acids, and esters [4], hydroxy-derivatives except 1-alkanols C_1–C_{10} [5], aromatic hydrocarbons [6], non-aromatic hydrocarbons (C_n, $n \geqslant 5$) except n-alkanes C_5–C_{16} [7], selected halogenated derivatives [8], and miscellaneous compounds [9].

2. DESCRIPTION OF DATABASE

2.1. Scope

Data for properties related to the state behavior of pure organic liquids such as density, specific and molar volume, relative properties (density and volume ratio, compression), volumetric coefficients (isobaric expansivity, isoentropic and isothermal compressibilities), and speed of sound are compiled and stored. The data for solids, vapors, and supercritical gases are also included in those cases when they come from the same source and experimental setup. Data compilation is not limited to direct experimental values but it also covers both the values generated from smoothing equations (if original direct experimental values are not available) and the values calculated from other properties (e.g., density at elevated pressures calculated from speed-of-sound data, data obtained from expansivities measured by calorimetric methods).

2.2. Database System

The system consists of several separate specialized components (Fig. 1):

- database of bibliographic information,
- database of experimental data,

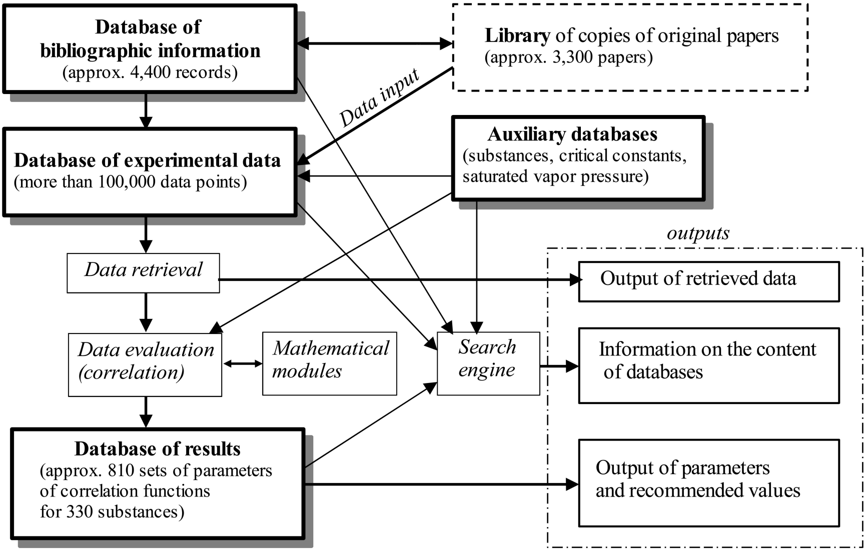

Fig. 1. Structure of database system and its content.

- database of results, and
- auxiliary databases.

The database of bibliographic information compiles references to literature sources. The content of each literature source is characterized in the database by property codes along with additional textual data. If a copy of the paper is available, then, using a reference code number, the record in the bibliographic database is linked to that copy stored in the library of papers (the reprint library). The library is a physical collection of copies of papers; some of the papers are also available as PDF files. The search program module (included in the "search engine," see Fig. 1) makes it feasible to extract the references that match a user-defined logical expression, to sort the retrieved references, and to create outputs (ASCII) in several fixed formats. The content of the bibliographic database is not limited to only sources that present experimental data but it also covers other papers related to state behavior (review papers that summarize thermodynamic properties, papers presenting equations of state, theoretical papers, etc.).

The core of the system is the database of experimental data that covers the original experimental data taken from literature sources. The description is presented in detail below.

Data for a particular substance, property, phase (solid, liquid, vapor, gas, and their combinations), and data type (pressure range: atmospheric, saturated, elevated and their combinations) can be retrieved from the database and stored on the disk in a temporary separate file. The content of the file may then be exported into an ASCII file with a fixed format for external use. Retrieved data can be correlated by a selected function of temperature or temperature and pressure (depending on the data type) using mathematical modules linked to the system. The results of each correlation can be stored in the database of results and exported into external ASCII files in various formats, including recommended values in a specified temperature or temperature and pressure range.

Auxiliary databases record characteristic information for each substance (substance data file: Chemical Abstracts Service Registry Number—CASRN, preferred name and alternative names, summary formula, molar mass), gas-liquid critical constants (critical temperature, pressure, and molar volume), and parameters of smoothing functions representing saturated vapor pressure (parameters of Antoine and/or Wagner equations). Data in these files can be used to distinguish among liquid, vapor, and supercritical gas for each particular T, p pair being entered into the database of experimental data. If the saturated vapor pressure is needed for a correlation (using the Tait equation for temperatures above the normal boiling point, see below) the values are calculated using the stored parameters.

The search engine is a set of programs that can be used to construct a number of various inquiries about the content of all databases and to output the required information.

2.3. Database of Experimental Data

The database of experimental state-behavior data (Fig. 2) is, along with its program package, designed to enter and store numerical and textual data compiled from the literature. Checks of logical data constraints and other precautions are built in the programs to minimize compiler's errors. The database is of the relational type consisting of several data files, each containing information of a particular character. The structure of the relations between individual data files ensures unique assignment of records of the data files in the hierarchy: literature reference → sample → data set header → experimental data (data set). Data for literature references are copied from the bibliographic database (Fig. 1) to the literature reference file (Fig. 2), and additional information is recorded (e.g., title of the paper, textual note). Each sample is defined by the CAS Registry Number (CASRN) of a substance and characterized by a description (source of the material, method of purification, final purity). Records in the file of data set headers define all information necessary to

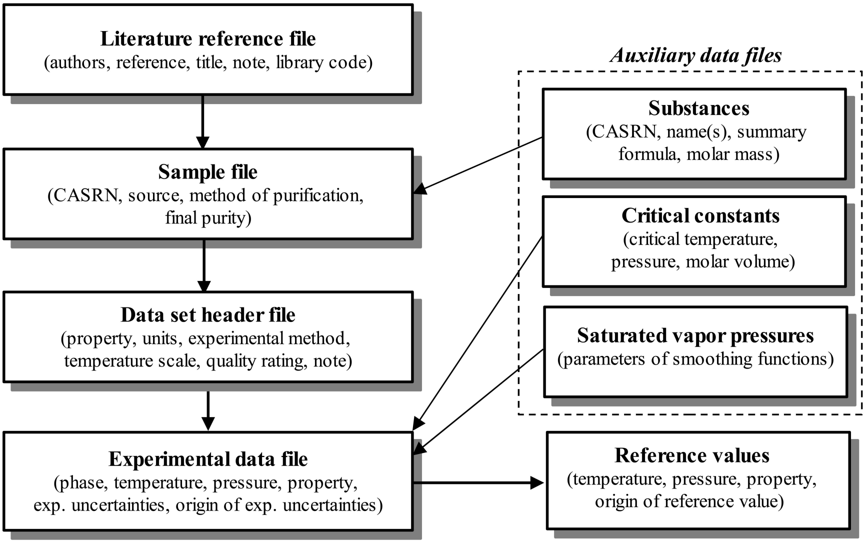

Fig. 2. Structure of database of experimental data.

characterize a set of experimental data (data set) for a particular sample. Each record consists of a code of property, a code of experimental method used, compiler's quality rating, data character (direct experimental, smoothed, calculated from other properties), temperature scale, and information on original units used for data presentation in the original source. Each record in the file of data set headers is related to one or more records in the file of experimental data (data set). Experimental data (values of temperature, pressure and property, their uncertainties, and a code of state, i.e., solid, liquid, vapor, or supercritical gas) are recorded in the file of experimental data. Experimental uncertainties are marked by a code of their origin, i.e., whether they were taken from the original data source or were estimated by a compiler. Each experimental value is entered with its unit and the number of digits as it is presented in the original source. The program recalculates the entered values into SI units (temperature into K, pressure into MPa, density, specific and molar volume into $\mathrm{kg \cdot m^{-3}}$, expansivity into $\mathrm{kK^{-1}}$, compressibilities into $\mathrm{GPa^{-1}}$, speed of sound into $\mathrm{m \cdot s^{-1}}$). Recalculated values along with information on numbers of digits in original values are stored in the file. Original values can be, therefore, restored, even with the original numbers of digits which is particularly useful when entered data are checked and corrected. Relative properties (density or volume ratio, compression) are also recalculated into density ($\mathrm{kg \cdot m^{-3}}$) if reference values are available. Reference values are recorded in a separate data file related to the file of experimental data.

3. DATA CORRELATION

External mathematical program modules that evaluate parameters of several functions of temperature or temperature and pressure for a selected set of retrieved data are linked to the database system. At present the following modules for correlation of density data are available.

(a) Densities at atmospheric pressure or at saturation (along the line of the liquid-vapor equilibrium) of substances for which the values of critical constants (critical temperature T_c, critical density d_c) are available are fitted by the function

$$d(T, \vec{a}) = d_c \left[1 + \sum_{i=0}^{N} a_i (1 - T_r)^{(i+1)/3} \right], \qquad T_r = T/T_c, \quad \vec{a} = \{a_i\}, \tag{1}$$

where $\{a_i\}$ are adjustable parameters and T absolute temperature. This function is preferably selected since it describes correctly density behavior in the vicinity of the critical point if $a_0 > 0$. For other substances the polynomial function,

$$d(T, \vec{a}) = \sum_{i=0}^{N} a_i (T/100)^i, \qquad \vec{a} = \{a_i\}, \tag{2}$$

where $\{a_i\}$ are adjustable parameters and T absolute temperature, is used. The polynomial function is used also for correlations of temperature dependences of other properties (volumetric coefficients, speed of sound) at atmospheric pressure or at saturation.

(b) Densities at elevated pressures (compressed liquid), i.e., density as a function of temperature and pressure, are correlated by the Tait equation,

$$d(T, p, \vec{c}, \vec{b}) = \frac{d(T, p_s(T))}{1 - C(T, \vec{c}) \ln \left\{ \dfrac{B(T, \vec{b}) + p}{B(T, \vec{b}) + p_s(T)} \right\}} \tag{3}$$

with

$$C(T, \vec{c}) = \sum_{i=0}^{N_C} c_i [(T - T_0)/100]^i, \qquad \vec{c} = \{c_i\}$$

$$B(T, \vec{b}) = \sum_{i=0}^{N_B} b_i [(T - T_0)/100]^i, \qquad \vec{b} = \{b_i\}$$

where $\{a_i\}$ and $\{b_i\}$ are adjustable parameters, T absolute temperature, and p pressure. The reference pressure p_s is the pressure selected as $p_s = 0.101325$ MPa at temperatures below the normal boiling temperature and $p_s = p(\text{saturated})$ for higher temperatures. Saturated vapor pressures ($p(\text{saturated})$) are calculated from smoothing functions (Antoine or Wagner equation) using parameters stored in the database of saturated vapor pressures. The parameter T_0 is a fixed pre-selected temperature, usually 298.15 K. The density $d(T, p_s(T))$ is the density of a liquid at temperature T and a reference pressure p_s.

Recently, a new module has been linked to the system that evaluates adjustable parameters of the function,

$$d(T, \vec{a}) = \sum_{i=0}^{N} \sum_{j=0}^{M} a_{ij} T^i p^j, \qquad \vec{a} = \{a_{ij}\}, \tag{4}$$

where $\{a_{ij}\}$ are adjustable parameters, T is absolute temperature, and p is pressure.

The Tait function (Eq. (3)) does not correlate compressed-liquid density $d(T, p)$ itself but represents the relative property $d(T, p)/d(T, p_s)$ where the reference density $d(T, p_s)$ is an input value. The values for the reference density reported by authors of compressed-liquid density data are preferably selected for correlations whenever possible. This minimizes significantly the effects of sample impurities and systematic experimental errors. Besides that, some experimental techniques output relative values (e.g., bellows piezometers that measure volume ratio $V(T, p)/V(T, p_s = 0.1\ \text{MPa})$ directly) and these values are correlated. If the reference value $d(T, p_s)$ is not available, then a value either extrapolated from data at elevated pressures is used (which is possible for isothermal data) or the value calculated from a selected $d(T)$ function is used. On the other hand, a selected reference line $d(T, p_s)$ (density of a particular sample or recommended data) can be employed when recommended density values $d(T, p)$ are generated using the Tait equation.

A weighted least-squares method is used to evaluate the parameters of correlation functions. The method is based on minimization of the objective function,

$$\phi(\vec{a}) = \sum_{j=1}^{N_\mathrm{p}} w_j [d_j - d(T_j, \vec{a})]^2 \qquad \text{or} \qquad \phi(\vec{a}) = \sum_{j=1}^{N_\mathrm{p}} w_j [d_j - d(T_j, p_j, \vec{a})]^2 \tag{5}$$

for Eqs. (1), (2), or (4), and

$$\phi(\vec{c}, \vec{b}) = \sum_{j=1}^{N_\mathrm{p}} w_j [d_j - d(T_j, P_j, \vec{c}, \vec{b})]^2 \tag{6}$$

for the Tait equation (Eq. (3)). Statistical weights w_j in Eqs. (5) and (6) are derived from experimental uncertainties recorded for each data point in the database of experimental data:

$$w_j = \omega_j / (\delta d_j)^2 \tag{7}$$

where δd_j is the experimental uncertainty (estimated experimental error) of the jth data point. The quantity ω_j is used to change the weights of individual data points by changing the value of ω_j (default setting is $\omega_j = 1$; for rejected data points, $\omega_j = 0$ holds) during the correlation procedure while original values of δd_j remain unchanged. Information relevant for adjusting the statistical weights (literature source of data, code of the reprint of the source in the reprint library, experimental method, sample purity, quality rating) is available on a computer screen. A user may thus take into account all recorded information and even go to the original source of data (copy of the paper) for more details. Built-in program modules allow checks of the shapes of dependences (extremes or inflection points, e.g., on $d(T)$ or $B(T)$ functions). The presence of extremes or inflection points may indicate an inconsistency of correlated data. Isothermal compressibility values calculated from the Tait equation for a particular temperature and at atmospheric pressure can also be displayed for each actual fit and a user may compare them with independent experimental data (direct experimental values or those calculated from speed-of-sound and heat-capacity data).

The external correlation program modules are controlled from the FoxPro environment. All settings as a selection of proper correlation function and the number of its adjustable parameters, adjustments of statistical weights, definition of temperature and pressure ranges, are performed in the FoxPro environment. Temporary ASCII data files are then exported and read by the external programs. After mathematical procedures are completed, the results are returned back into the FoxPro environment and displayed. The process can be repeated until a final fit is reached.

Results of each final correlation are stored in the database of results and may either be added into the database as a new record or previous results can be overwritten. The stored data are values of adjustable parameters, values of external parameters (if applicable, e.g., critical temperature, critical density), temperature and pressure ranges, statistical characteristics of the fit (root-mean-square deviations and weighted standard deviations of the fit and of individual data sets) and a correlation protocol. The protocol is the list of all data used for the correlation with the setting of statistical weights. The protocol enables one to employ the stored setting when the correlation is updated with new data. A covariance matrix of each fit is also stored which may be employed for calculation of standard deviation estimates of the values calculated from the correlation function.

4. CONCLUSION

The present system proved to be an efficient tool for collection and evaluation of state behavior data. A significant amount of published experimental data on the state behavior of pure organic liquids has been compiled, and the content of the database of experimental data is being currently updated with newly found and published data. The database itself in not available to public, the results of the data evaluation has been published [1–9] for 315 pure organic liquids.

ACKNOWLEDGMENTS

This work was supported by the Czech Ministry of Education, Youth and Sports (fund MSM223400008 and Grants ME 312 and ME 514).

REFERENCES

1. I. Cibulka, *Fluid Phase Equilib.* **89**:1 (1993).
2. I. Cibulka and M. Ziková, *J. Chem. Eng. Data* **39**:876 (1994).
3. I. Cibulka and L. Hnědkovský, *J. Chem. Eng. Data* **41**:657 (1996).
4. I. Cibulka, L. Hnědkovský, and T. Takagi, *J. Chem. Eng. Data* **42**:2 (1997).
5. I. Cibulka, L. Hnědkovský, and T. Takagi, *J. Chem. Eng. Data* **42**:415 (1997).
6. I. Cibulka and T. Takagi, *J. Chem. Eng. Data* **44**:411 (1999).
7. I. Cibulka and T. Takagi, *J. Chem. Eng. Data* **44**:1105 (1999).
8. I. Cibulka, T. Takagi, and K. Růžička, *J. Chem. Eng. Data* **46**:2 (2001).
9. I. Cibulka and T. Takagi, *J. Chem. Eng. Data* **47**:1037 (2002).