Int J Thermophys (2007) 28:824–834
DOI 10.1007/s10765-007-0224-9

Rapid Evaluation of Prediction Methods with DIPPR's Automated Property Prediction Package

J. R. Rowley · W. V. Wilding · J. L. Oscarson · R. L. Rowley

Published online: 16 August 2007

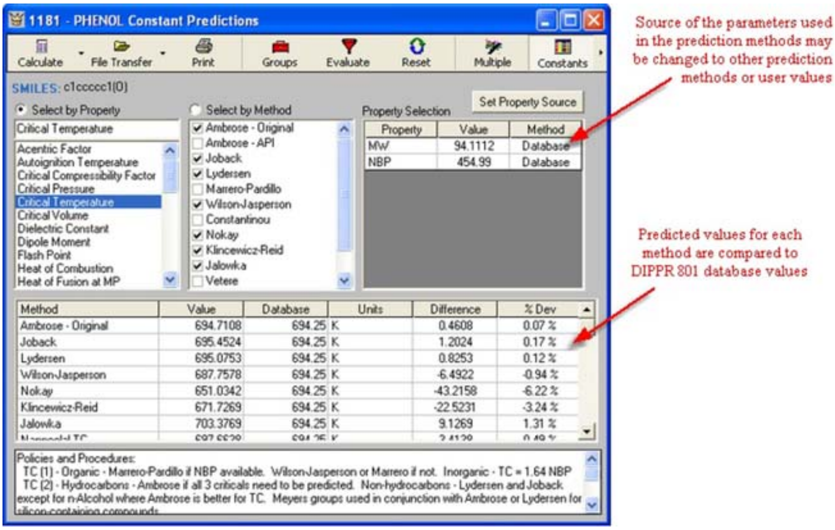
Abstract An automated property prediction package has been developed that permits rapid evaluation of group-contribution, corresponding states, empirical, and theoretical property estimation methods. The property prediction package, which is part of the DIPPR® Information And Data Evaluation Manager (DIADEM) software, is used in conjunction with the DIPPR® 801 database to develop and test new prediction methods. The software is freely available to all DIPPR sponsor companies, but is also commercially available. The estimation engine is based on an automated SMILES (Simplified Molecular Input Line Entry Specification) formula parser to provide required molecular structural information, retrieval of required secondary properties from the DIPPR® database, and defined rules for the method. Automatic comparisons of predicted values to experimental data in the DIPPR® database can be made for properties at specified accuracy levels, by chemical family or type, or over the entire database. This allows evaluation of the relative effectiveness of methods for specific chemical families and tailoring of the selected method to specific chemical classes. New methods can readily be added by input using a simple input form. Nearly 200 thermophysical property prediction methods are currently available in DIADEM.

Keywords Automated · DIPPR · Estimation · Prediction · Property · SMILES

1 Introduction

The DIPPR 801 database is an important resource for process design and development, providing pure-component property values for 29 constant properties and 15

J. R. Rowley (✉) · W. V. Wilding · J. L. Oscarson · R. L. Rowley
DIPPR Thermophysical Property Laboratory and Chemical Engineering Department, Brigham Young University, Provo, UT 84602, USA
e-mail: jrr46@byu.net



SMILES: c1ccccc1(O)

● Select by Property

Critical Temperature

Acentric Factor
Autoignition Temperature
Critical Compressibility Factor
Critical Pressure
Critical Temperature
Critical Volume
Dielectric Constant
Dipole Moment
Flash Point
Heat of Combustion
Heat of Fusion at MP

○ Select by Method

☑ Ambrose - Original
☐ Ambrose - API
☑ Joback
☑ Lydersen
☐ Marrero-Pardillo
☑ Wilson-Jasperson
☐ Constantinou
☑ Nokay
☑ Klincewicz-Reid
☑ Jalowka
☐ Vetere

Property Selection

Property	Value	Method
MW	94.1112	Database
NBP	454.99	Database

Method	Value	Database	Units	Difference	% Dev
Ambrose - Original	694.7108	694.25	K	0.4608	0.07 %
Joback	695.4524	694.25	K	1.2024	0.17 %
Lydersen	695.0753	694.25	K	0.8253	0.12 %
Wilson-Jasperson	687.7578	694.25	K	-6.4922	-0.94 %
Nokay	651.0342	694.25	K	-43.2158	-6.22 %
Klincewicz-Reid	671.7269	694.25	K	-22.5231	-3.24 %
Jalowka	703.3769	694.25	K	9.1269	1.31 %

Policies and Procedures:
TC (1) - Organic - Marrero-Pardillo if NBP available. Wilson-Jasperson or Marrero if not. Inorganic - TC = 1.64 NBP
TC (2) - Hydrocarbons - Ambrose if all 3 criticals need to be predicted. Non-hydrocarbons - Lydersen and Joback except for n-Alcohol where Ambrose is better for TC. Meyers groups used in conjunction with Ambrose or Lydersen for silicon-containing compounds.

Source of the parameters used in the prediction methods may be changed to other prediction methods or user values

Predicted values for each method are compared to DIPPR 801 database values

Fig. 1 DIADEM's prediction package interface for a single component

temperature-dependent properties for approximately 2,000 compounds of industrial importance. Four foundational principles have guided the efforts on the database project and have been key to its success. These are (a) industrial sponsor control, (b) critical evaluation, (c) consistency, and (d) completeness.

The fourth principle, completeness, requires that whenever possible, values for all properties are provided. When experimental data are lacking, the principle of completeness requires that prediction techniques used to supply the needed values.

Knowing which prediction method to use, however, can be difficult. It is important to understand the limitations of a prediction method, and to which compounds the method may be applied. BYU-DIPPR® 801 has created a prediction package as part of the DIPPR® Information And Data Evaluation Manager (DIADEM) software that enables rapid property estimation and evaluation of group-contribution, corresponding states, empirical, and theoretical property estimation methods.

2 Prediction Package Overview

DIADEM's prediction package contains simple user interfaces to estimate properties for a single compound (Fig. 1), to input new prediction methods to the software, and to compare the performance of existing methods. Currently, DIADEM has nearly 200 methods divided into two main classes: non-structural and group-contribution methods.

2.1 Non-structural Prediction Methods

Non-structural are those methods that estimate a property without knowledge pertaining to the family or structure of the compounds. These methods usually employ a correlation between the property to be estimated and one or more of the compound's physical properties. In DIADEM, the data needed for property predictions are automatically extracted from the DIPPR® 801 database if available. If the supporting property values required for the predictions are not available, DIADEM predicts those also using DIPPR® 801 primary methods. Users may also override the default method and instruct DIADEM to obtain data from another prediction method, or use user-input values.

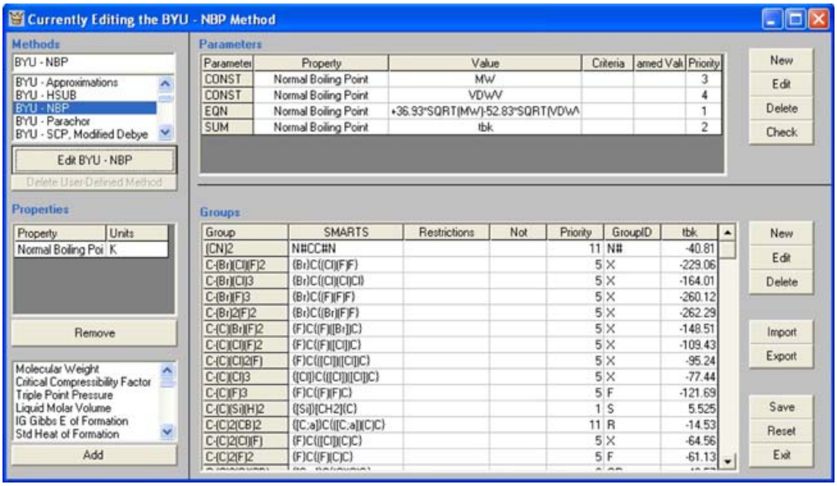
2.2 Group-Contribution Prediction Methods

Group-contribution methods are those methods that utilize knowledge of the compound's structure to estimate a property. Group-contribution methods range from using different parameters or coefficients for different classes or families of compounds, to utilizing weighted contributions of individual atoms or functional groups. In order to perform group-contribution property estimations, DIADEM first uses Simplified Molecular Input Line Entry Specification (SMILES) formulas for the compound to parse the molecular structure into atomic contributions. The method's contributions are then extracted from a database where each contribution has been identified using a modified Smiles Arbitrary Target Specification (SMARTS) notation. Each contribution is compared with segments of the molecule reconstructed from forward parsing of the atomic contributions. If a match between the method contributions and portions of the molecule is made, the participating atoms from the molecule are recorded as matched to avoid duplicate use of the same atoms.

In order to ensure proper matching of group contributions, each contribution from the method is assigned a numerical priority according to its relative size and detail. Contributions with higher priority, and thus more detailed specifications, are utilized before lesser-specified contributions; i.e., CH_3 would be used before CH, an acid group would be used before a carbonyl, etc. This prioritization of groups creates an easy way to adjust the order of the groups that DIADEM finds, and thus ensures accurate representation of the molecule. In cases when the user disagrees with DIADEM's interpretation of the priority of an author's groups, users may easily switch the groups DIADEM employs for the compound and re-evaluate the property with the user-specified groups.

2.3 Addition of New Methods

DIADEM's prediction package is setup to allow continual addition of new methods and modification of existing methods. All supporting data for the methods are stored in a database, and the code is designed to be non-method specific, so that updating, modifying, and adding new methods is simple. Thus, users may implement and



Parameter	Property	Value	Criteria	amed Val	Priority
CONST	Normal Boiling Point	MW			3
CONST	Normal Boiling Point	VDWV			4
EQN	Normal Boiling Point	+36.93*SQRT(MW)-52.83*SQRT(VDW\			1
SUM	Normal Boiling Point	tbk			2

Property	Units
Normal Boiling Poi	K

Group	SMARTS	Restrictions	Not	Priority	GroupID	tbk
(CN)2	N#CC#N			11	N#	-40.81
C-(Br)(Cl)(F)2	(Br)C((Cl)(F)F)			5	X	-229.06
C-(Br)(Cl)3	(Br)C((Cl)(Cl)Cl)			5	X	-164.01
C-(Br)(F)3	(Br)C((F)(F)F)			5	X	-260.12
C-(Br)2(F)2	(Br)C((Br)(F)F)			5	X	-262.29
C-(C)(Br)(F)2	(F)C((F)([Br])C)			5	X	-148.51
C-(C)(Cl)(F)2	(F)C((F)([Cl])C)			5	X	-109.43
C-(C)(Cl)2(F)	(F)C(([Cl])([Cl])C)			5	X	-95.24
C-(C)(Cl)3	([Cl])C(([Cl])([Cl])C)			5	X	-77.44
C-(C)(F)3	(F)C((F)(F)C)			5	F	-121.69
C-(C)(Si)(H)2	([Si])[CH2](C)			1	S	5.525
C-(C)2(CB)2	([C;a])C(([C;a])(C)C)			11	R	-14.53
C-(C)2(Cl)(F)	(F)C(([Cl])(C)C)			5	X	-64.56
C-(C)2(F)2	(F)C((F)(C)C)			5	F	-61.13

Fig. 2 User-interface to add or edit prediction methods. Groups are first defined using a modified SMARTS formula, then assigned a priority by size and/or detail of the group

evaluate any number of methods in addition to the nearly 200 methods supplied with DIADEM.

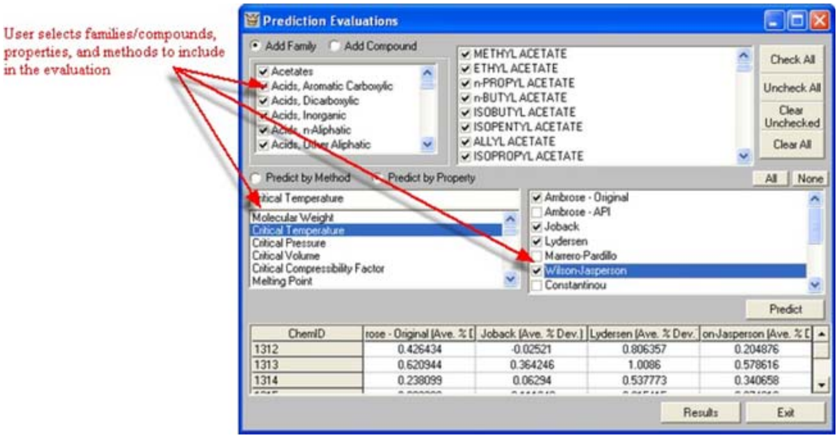
Methods are added and edited using a simple user interface (Fig. 2). Users input the parameters, correlations, and group contributions (if applicable) with their corresponding priorities. Groups are added by inputting a SMARTS formula that describes each group, and assigning a relative priority to each group. Therefore, knowledge of SMARTS is necessary to add group-contribution methods; however, detailed tutorials are available on the web for those unfamiliar with SMARTS [1].

3 Method Evaluation

3.1 Evaluator Overview

DIADEM utilizes a simple user-interface to perform an evaluation of a single prediction method, or compare the performance of multiple methods (Fig. 3). Using each of the selected methods, DIADEM will predict the applicable property. Predicted values are then compared to “accepted” (DIPPR 801 uses the classification of accepted for the value recommended by the DIPPR 801 staff) experimental data within the DIPPR® 801 database to determine an average absolute percent deviation. Users may opt to evaluate methods by comparing predicted values to selected compounds, entire families or classes of compounds, or using the entire database.

When the DIADEM evaluation is complete, the methods are ordered by lowest overall average absolute deviation (AAD) from experimental data. DIADEM displays the AAD for each method, the compounds that resulted in the lowest and highest deviation for each method, and a list of deviations for each method sorted by compound family (Fig. 4). Additionally, the evaluation results are stored for later access.


User selects families/compounds, properties, and methods to include in the evaluation

☑ Acetates
☑ Acids, Aromatic Carboxylic
☑ Acids, Dicarboxylic
☑ Acids, Inorganic
☑ Acids, n-Aliphatic
☑ Acids, Other Aliphatic

☑ METHYL ACETATE
☑ ETHYL ACETATE
☑ n-PROPYL ACETATE
☑ n-BUTYL ACETATE
☑ ISOBUTYL ACETATE
☑ ISOPENTYL ACETATE
☑ ALLYL ACETATE
☑ ISOPROPYL ACETATE

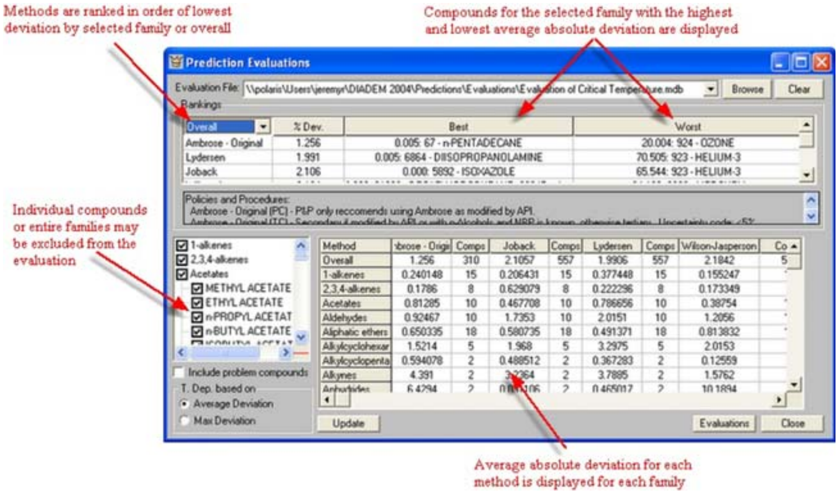
itical Temperature

Molecular Weight
Critical Temperature
Critical Pressure
Critical Volume
Critical Compressibility Factor
Melting Point

☑ Ambrose - Original
□ Ambrose - API
☑ Joback
☑ Lydersen
□ Marrero-Pardillo
☑ Wilson-Jasperson
□ Constantinou

ChemID	rose - Original (Ave. % D	Joback (Ave. % Dev.)	Lydersen (Ave. % Dev.	on-Jasperson (Ave. % D
1312	0.426434	-0.02521	0.806357	0.204876
1313	0.620944	0.364246	1.0086	0.578616
1314	0.238099	0.06294	0.537773	0.340658

Fig. 3 Setup of critical temperature evaluation for Ambrose [2,3], Joback [4], Lydersen [5], and Wilson-Jasperson [6] methods


Methods are ranked in order of lowest deviation by selected family or overall

Compounds for the selected family with the highest and lowest average absolute deviation are displayed

Individual compounds or entire families may be excluded from the evaluation

Average absolute deviation for each method is displayed for each family

Evaluation File: \\polaris\Users\jeremyr\DIADEM 2004\Predictions\Evaluations\Evaluation of Critical Temperature.mdb

Rankings

Overall	% Dev.	Best	Worst
Ambrose - Original	1.256	0.005: 67 - n-PENTADECANE	20.004: 924 - OZONE
Lydersen	1.991	0.005: 6864 - DIISOPROPANOLAMINE	70.505: 923 - HELIUM-3
Joback	2.106	0.000: 5892 - ISOXAZOLE	65.544: 923 - HELIUM-3

Policies and Procedures:
Ambrose - Original (PC) - P&P only reccomends using Ambrose as modified by API.

☑ 1-alkenes
☑ 2,3,4-alkenes
☑ Acetates
☑ METHYL ACETATE
☑ ETHYL ACETATE
☑ n-PROPYL ACETAT
☑ n-BUTYL ACETATE

□ Include problem compounds

T. Dep. based on
● Average Deviation
○ Max Deviation

Method	ıbrose - Origi	Comps	Joback	Comps	Lydersen	Comps	Wilson-Jasperson	Co
Overall	1.256	310	2.1057	557	1.9906	557	2.1842	5
1-alkenes	0.240148	15	0.206431	15	0.377448	15	0.155247	
2,3,4-alkenes	0.1786	8	0.629079	8	0.222296	8	0.173349	
Acetates	0.81285	10	0.467708	10	0.786656	10	0.38754	
Aldehydes	0.92467	10	1.7353	10	2.0151	10	1.2056	
Aliphatic ethers	0.650335	18	0.580735	18	0.491371	18	0.813832	
Alkylcyclohexar	1.5214	5	1.968	5	3.2975	5	2.0153	
Alkylcyclopenta	0.594078	2	0.488512	2	0.367283	2	0.12559	
Alkynes	4.391	2	3.2364	2	3.7885	2	1.5762	
Anhydrides	6.4294	2	[illegible]	2	0.465017	2	10.1894	

Fig. 4 DIADEM's prediction package evaluation results interface

Prediction methods for two properties were evaluated for this work. Shown in Tables 1 and 2 are the AAD results by family for the critical temperature and normal boiling point.

3.2 Benefits of Prediction Method Evaluation

Evaluation of methods in this manner provides several benefits. From the results it is easy to determine which prediction methods apply to specific families. Extensive evaluation of properties by chemical families aids the user in selecting the most accurate

Table 1 Evaluation results for Ambrose, Joback, Lydersen, and Wilson-Jasperson methods of predicting critical temperature

Family	Comps	AAD (%)			
		Ambrose	Joback	Lydersen	Wilson-Jasperson
Overall	565	4.06	3.49	3.38	3.57
1-alkenes	15	0.24	0.21	0.38	0.16
2,3,4-alkenes	8	0.18	0.63	0.22	0.17
Acetates	10	0.81	0.47	0.79	0.39
Aldehydes	10	0.92	1.74	2.02	1.21
Aliphatic ethers	18	0.65	0.58	0.49	0.81
Alkylcyclohexanes	5	1.52	1.97	3.30	2.02
Alkylcyclopentanes	2	0.59	0.49	0.37	0.13
Alkynes	3	36.26	35.49	35.86	34.38
Anhydrides	2	6.43	0.03	0.47	10.19
Aromatic alcohols	17	0.66	1.42	1.43	0.96
Aromatic amines	13	1.01	0.97	1.06	1.03
Aromatic chlorides	3	0.60	0.55	0.58	0.45
C, H, Br compounds	5	0.96	1.82	1.10	1.42
C, H, F compounds	43	6.21	1.17	1.71	1.76
C, H, I compounds	3	0.59	0.51	0.49	1.20
C, H, multihalogen compounds	21	8.19	0.63	0.56	0.82
C, H, NO_2 compounds	1	0.02	0.01	0.02	0.11
C1/C2 aliphatic chlorides	12	0.59	0.89	0.76	0.52
C3 & higher aliphatic chlorides	6	0.72	0.97	0.91	0.71
Cycloaliphatic alcohols	1	1.17	4.21	4.73	1.56
Cycloalkanes	4	0.52	0.45	0.33	1.93
Cycloalkenes	2	0.22	0.07	0.14	0.66
Dialkenes	3	0.78	0.57	0.89	0.79
Dimethylalkanes	11	1.24	0.43	0.39	1.37
Diphenyl/polyaromatics	4	0.89	1.02	1.01	1.55
Elements	20	23.60	21.99	22.34	17.35
Epoxides	6	1.07	1.01	0.88	2.10
Formates	5	0.90	1.52	0.70	2.68
Inorganic acids	5	2.99	5.72	4.72	3.37
Inorganic bases	1	0.15	2.62	1.02	0.07
Inorganic gases	25	17.59	16.19	15.20	15.87
Inorganic halides	11	15.87	13.08	12.71	11.71
Isocyanates/diisocyanates	1	0.54	3.36	1.82	12.14
Ketones	25	0.92	1.30	1.46	8.53
Mercaptans	5	0.17	0.37	0.46	0.17
Methylalkanes	11	1.49	0.35	0.41	0.53
Methylalkenes	4	0.50	1.37	2.20	0.65

Table 1 continued

Family	Comps	AAD (%)			
		Ambrose	Joback	Lydersen	Wilson-Jasperson
Multiring cycloalkanes	2	1.62	1.04	3.00	0.76
n-alcohols	15	0.37	2.48	2.66	0.97
n-aliphatic acids	8	1.02	1.30	1.02	1.26
n-aliphatic primary amines	4	0.84	0.58	0.69	0.74
n-alkanes	5	0.07	0.64	1.11	0.19
n-alkylbenzenes	1	1.05	0.37	0.46	0.76
Naphthalenes	2	1.39	1.18	1.00	1.19
Nitriles	8	3.34	1.88	2.16	2.01
Organic/inorganic compounds	1	2.16	2.86	2.42	7.52
Other[a] aliphatic acids	3	1.64	1.50	1.61	1.02
Other aliphatic alcohols	25	1.38	2.24	2.35	0.85
Other aliphatic amines	12	0.89	0.72	0.80	1.23
Other alkanes	18	1.04	0.53	0.60	2.66
Other alkylbenzenes	9	0.99	0.98	0.89	1.05
Other amines, imines	3	0.38	0.32	0.73	1.25
Other condensed rings	1	0.87	0.82	1.06	5.00
Other ethers/diethers	8	0.81	0.51	0.63	1.30
Other hydrocarbon rings	1	0.53	0.78	0.35	1.82
Other inorganics	5	3.01	8.02	6.72	2.86
Other polyfunctional C, H, O	8	1.56	3.92	4.36	2.69
Other saturated aliphatic esters	5	1.47	1.42	1.57	1.95
Polyfunctional amides/amines	5	8.45	2.44	2.54	3.72
Polyfunctional C, H, O, halide	11	4.35	2.82	2.79	4.74
Polyfunctional C, H, O, N	2	0.06	0.99	0.86	4.74
Polyfunctional C, H, O, S	1	1.13	1.67	1.76	9.62
Polyfunctional esters	6	0.75	0.84	0.83	1.63
Polyols	7	5.43	8.72	8.96	7.08
Propionates and butyrates	11	0.41	0.51	0.57	0.45
Silanes/siloxanes	23	5.72	8.70	5.29	5.44
Sulfides/thiophenes	12	1.80	1.54	1.20	1.58
Terpenes	2	0.96	1.59	1.40	0.64

[a] The term "other" in Tables 1 and 2 refer to those compounds that are not accurately categorized by the more detailed classification. For example, "other polyfunctional C, H, O" are compounds that contain C, H, and O atoms, but can not be readily classified in any of the other polyfunctional families

Table 2 Evaluation results for BYU-NBP [7], Joback [4], Miller [8], and Stein [9] methods of predicting normal boiling point

		AAD (%)			
Family	Comps	BYU-NBP	Joback	Miller	Stein
Overall	1566	12.53	17.20	10.19	15.59
1-alkenes	31	2.89	5.35	3.36	4.15
2,3,4-alkenes	28	1.37	2.60	2.10	3.42
Acetates	24	1.39	1.57	3.19	0.68
Aldehydes	29	1.30	3.20	6.46	1.77
Aliphatic ethers	35	1.87	2.39	2.85	1.80
Alkylcyclohexanes	18	1.51	1.12	5.25	3.52
Alkylcyclopentanes	11	1.22	1.48	3.69	2.65
Alkynes	17	1.29	6.85	3.66	3.39
Anhydrides	8	3.56	4.87	5.02	4.23
Aromatic alcohols	35	2.50	4.22	12.52	1.43
Aromatic amines	30	6.92	3.27	6.93	2.66
Aromatic carboxylic acids	6	6.86	4.65	9.25	1.43
Aromatic chlorides	17	0.89	2.06	6.87	1.76
Aromatic esters	19	3.25	9.10	4.77	1.89
C, H, Br compounds	18	2.79	2.55	12.08	2.48
C, H, F compounds	49	18.99	9.30	7.68	19.74
C, H, I compounds	13	1.92	3.42	8.12	2.95
C, H, multihalogen compounds	37	4.74	10.34	4.30	13.88
C, H, NO_2 compounds	11	1.12	14.23	5.69	8.09
C1/C2 aliphatic chlorides	20	2.15	3.56	5.23	7.08
C3 & higher aliphatic chlorides	31	1.23	7.80	4.46	3.89
Cycloaliphatic alcohols	10	9.12	5.31	4.78	2.44
Cycloalkanes	6	4.20	5.37	1.25	5.53
Cycloalkenes	10	2.33	1.88	4.41	2.81
Dialkenes	28	1.50	2.48	4.12	2.83
Dicarboxylic acids	1	1.31	5.74	6.14	2.96
Dimethylalkanes	21	1.49	1.67	2.02	4.40
Diphenyl/polyaromatics	12	3.01	9.31	7.21	2.89
Elements	30	312.73	523.90	43.84	490.67
Epoxides	14	6.67	2.31	4.56	3.34
Ethyl & higher alkenes	12	1.05	1.76	3.16	2.20
Formates	12	0.67	1.65	3.68	0.91
Inorganic acids	9	32.53	22.13	47.33	19.83
Inorganic bases	3	37.07	37.93	49.02	34.58
Inorganic gases	26	54.43	43.15	25.61	47.08
Inorganic halides	22	41.50	36.59	49.02	37.22
Isocyanates/diisocyanates	5	0.68	19.85	2.26	4.19

Table 2 continued

Family	Comps	AAD (%)			
		BYU-NBP	Joback	Miller	Stein
Ketones	41	2.23	3.22	3.55	2.75
Mercaptans	21	1.52	3.89	4.94	2.35
Methylalkanes	18	1.69	2.78	1.88	3.07
Methylalkenes	22	1.27	2.29	3.59	1.96
Multiring cycloalkanes	3	2.33	1.04	5.55	5.42
n-alcohols	12	4.49	2.91	9.52	1.50
n-aliphatic acids	26	2.98	3.27	8.89	1.07
n-aliphatic primary amines	13	1.84	3.48	7.08	2.09
n-alkanes	20	2.73	10.12	2.58	8.40
n-alkylbenzenes	10	0.91	0.80	2.77	1.97
Naphthalenes	13	1.36	2.19	3.13	1.56
Nitriles	24	2.73	12.86	20.94	6.70
Organic salts	16	25.13	16.06	48.87	8.18
Organic/inorganic compounds	7	24.10	20.58	39.96	18.48
Other[a] aliphatic acids	16	1.73	3.51	5.19	0.98
Other aliphatic alcohols	39	2.05	4.55	9.15	1.42
Other aliphatic amines	22	1.98	6.05	6.22	2.56
Other alkanes	23	2.09	1.92	1.98	6.98
Other alkylbenzenes	49	0.99	2.80	3.49	1.65
Other amines, imines	33	4.79	5.98	9.02	3.15
Other condensed rings	11	1.17	1.94	12.57	3.47
Other ethers/diethers	20	4.41	3.32	5.36	2.65
Other hydrocarbon rings	11	4.75	1.63	3.53	3.02
Other inorganic salts	1	19.61	20.07	99.99	21.67
Other inorganics	12	58.08	53.31	63.31	53.76
Other monoaromatics	15	1.24	2.81	6.06	1.06
Other polyfunctional C, H, O	42	2.51	5.11	10.92	2.93
Other polyfunctional organics	5	6.42	8.40	30.92	4.16
Other saturated aliphatic esters	18	5.40	9.65	5.05	2.46
Peroxides	4	1.84	5.68	12.38	11.13
Polyfunctional acids	3	15.88	7.94	9.25	4.62
Polyfunctional amides/amines	22	12.69	12.13	13.87	5.08
Polyfunctional C, H, N, halide, (o)	9	9.78	10.58	7.16	6.37
Polyfunctional C, H, O, halide	45	5.15	7.81	9.60	8.35
Polyfunctional C, H, O, N	20	6.40	8.08	10.22	3.04
Polyfunctional C, H, O, S	9	8.00	17.26	11.63	6.18
Polyfunctional esters	20	3.01	5.61	7.80	2.41
Polyfunctional nitriles	2	2.67	0.98	11.11	6.50
Polyols	32	4.01	6.89	27.79	4.49

Table 2 continued

Family	Comps	AAD (%) BYU-NBP	Joback	Miller	Stein
Propionates and butyrates	13	0.58	1.15	2.07	0.84
Silanes/siloxanes	43	18.44	11.12	8.52	6.53
Sodium salts	7	61.81	66.11	77.30	68.15
Sulfides/thiophenes	41	4.56	3.04	4.35	3.19
Terpenes	7	1.60	1.30	5.66	2.25
Unsaturated aliphatic esters	18	1.85	2.49	2.80	1.07

[a] The term "other" in Tables 1 and 2 refer to those compounds that are not accurately categorized by the more detailed classification. For example, "other polyfunctional C, H, O" are compounds that contain C, H, and O atoms, but can not be readily classified in any of the other polyfunctional families

method for the compound of interest. This reduces time required in literature research and increases overall accuracy of the estimated properties.

Knowledge that a particular family produces higher deviations for a particular method also indicates that functional groups unique to that family may be poorly determined and need tuning. Work may then be focused on acquiring or modifying contributions for just those specific groups with higher deviations, rather than creating an entirely new method. Additionally, consistently high deviations using several methods may indicate families and/or properties in need of further experimental data.

4 Conclusions

DIADEM's prediction method evaluator is a powerful tool for evaluating prediction techniques and assisting in the improvement of existing techniques and development of new techniques. Evaluating methods using the whole DIPPR 801 database and comparing results for specific families of compounds pinpoints functional groups that existing methods do not accurately represent. This guides the wise use of existing methods and focuses correlational and experimental efforts for efficient improvements. The modular form of the way in which new methods can be entered allows users to input, test, and further develop new prediction methods in real time.

References

1. http://www.daylight.com, Daylight Chemical Information Systems, Inc. (2005)
2. D. Ambrose, *Vapor—Liquid Critical Properties* (National Physical Laboratory Report Chem 107, NPL, Middlesex, United Kingdom, 1980)
3. D. Ambrose, *Correlation and Estimation of Vapour–Liquid Critical Properties: I. Critical Temperatures of Organic Compounds* (National Physical Laboratory Report Chem. 92, NPL, Middlesex, United Kingdom, 1978)
4. K.G. Joback, M. S. Thesis in Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts (1984)
5. A.L. Lydersen, *Estimation of Critical Properties of Organic Compounds* (University of Wisconsin Coll. Eng. Exp. Stn. Rep. 3, Madison, Wisconsin, 1955)

6. G.M. Wilson, L.V. Jasperson, *Critical Constants T_C, P_C, Estimation based on Zero, First and Second Order Methods* (AIChE Spring Meeting, New Orleans, Louisiana, 1996)
7. D. Ericksen, W.V. Wilding, J.L. Oscarson, R.L. Rowley, J. Chem. Eng. Data **47**, 1293 (2002)
8. W.J. Lyman, W.F. Reehl, D.H. Rosenblatt, *Handbook of Chemical Property Estimation Methods* (McGraw-Hill, New York, 1982)
9. S.E. Stein, R.L. Brown, J. Chem. Inf. Comp. Sci. **34**, 581 (1994)