Boiling Points of Halogenated Aliphatic Compounds: A Quantitative Structure−Property Relationship for Prediction and Validation

Tomas Öberg*

Department of Biology and Environmental Science, University of Kalmar, SE-391 82 Kalmar, Sweden

Received August 22, 2003

Halogenated aliphatic compounds have many technical uses, but substances within this group are also ubiquitous environmental pollutants that can affect the ozone layer and contribute to global warming. The establishment of quantitative structure−property relationships is of interest not only to fill in gaps in the available database but also to validate experimental data already acquired. The three-dimensional structures of 240 compounds were modeled with molecular mechanics prior to the generation of empirical descriptors. Two bilinear projection methods, principal component analysis (PCA) and partial-least-squares regression (PLSR), were used to identify outliers. PLSR was subsequently used to build a multivariate calibration model by extracting the latent variables that describe most of the covariation between the molecular structure and the boiling point. Boiling points were also estimated with an extension of the group contribution method of Stein and Brown.

INTRODUCTION

The boiling point of a compound is an important property, and like the vapor pressure it provides an indication of the attractive forces between the molecules. These intermolecular forces are directly related to the structure of the compound, and hence the boiling point may be correlated to the structure.[1] The boiling point is used to identify compounds, to assess their volatility, and also to estimate a number of other physical properties. Experimental data are available for a large number of volatile compounds, but estimation methods are of interest since they provide the means to fill in gaps in databases and can be used for validation of data.[2]

Halogenated aliphatic compounds are used as refrigerants, blowing agents, anesthetics, solvents, and fire extinguishers, etc. Many of the uses depend on the high strength of the fluorine−carbon and chlorine−carbon bonds, which make most of these compounds extremely stable. The nonreactivity and volatility of short-chained halogenated compounds, together with large production volumes, are the background for concern regarding ozone depletion and greenhouse warming potential.[3,4] The Montreal Protocol currently controls 96 halogenated aliphatics (C_1–C_3) damaging to the ozone layer, and in addition both hydrofluorocarbons and perfluorocarbons are controlled by the Kyoto Protocol to the Convention on Climate Change.[5,6]

The boiling point is of particular interest for these volatile and environmentally relevant compounds. Experimentally determined boiling points have been published for many of them, but there are also gaps in the available data. It is therefore of great interest that several reports have appeared regarding the estimation of the boiling point for short-chained halogenated aliphatic compounds and derivatives, either specifically or as part of a larger set of compounds.[2,7−18] The development of estimation methods is, however, always limited by the availability of reliable calibration data.

The publication in 2001 of three decades of experimental measurements on halogenated organic compounds at Imperial Chemical Industries (ICI) was a major addition to the available database.[19] The chemicals included several groups of compounds, notably halogenated aliphatic hydrocarbons and ethers, as well as halogenated cyclic compounds. The largest group was the halogenated aliphatic compounds. The author of the article on the ICI study also noted that the assessment and critical evaluation of boiling points of halogenated organic compounds is very limited and inadequate, and further stated the difficulties in relating measured values to molecular structures.

Questions of accuracy and precision are a common concern regarding experimental measurements, and several authors have addressed this issue with regard to boiling point determinations.[2,13,16,20,21] A general conclusion that can be drawn from the data sets discussed in these reports is that the experimental error will fall within the interval 5−10 K.

The literature contains many estimation methods for the boiling point, and the additive group contribution methods are well-known.[22−24] These methods are fairly robust and can be applied to a wide variety of organic molecules; however, a robust and general method will by definition lack some accuracy and precision when considering local phenomena. Topological, geometrical, and electronic descriptors can give more in-depth descriptions of the molecules and serve as a basis for developing predictive models with improved accuracy and precision.[20,25−27]

Quantitative structure−property relationships (QSPRs) rely on a basic assumption of similarity; i.e., the models are only valid for similar compounds. Outlier detection capabilities and adherence to the parsimony principle are both critical to the future reliability of any multivariate calibration model.[28] Linear regression has traditionally been the method of choice for QSPR modeling, but many of the recent studies on boiling point estimation rely on neural networks.[2,8,12,18,20,29−32] Lack of interpretability is a major drawback with the more complex neural network modeling

* Phone: 46 480-44 73 00; fax; 46 480-44 73 05; e-mail: tomas.oberg@hik.se.

Table 1. Number of Compounds with the Different Elemental Compositions

atom type	for given no. of atoms										
	0	1	2	3	4	5	6	7	8	9	>9
fluorine	15	17	70	52	37	30	6	5	4	1	3
chlorine	79	75	55	21	10	0	0	0	0	0	0
bromine	148	65	23	4	0	0	0	0	0	0	0
carbon	0	15	71	128	22	2	0	2	0	0	0
hydrogen	33	59	49	42	31	13	7	3	3	0	0

procedure, and performance has also been questioned.[33,34] Furthermore, neither traditional linear regression nor the neural networks incorporate outlier detection as a standard feature. Projection-based methods such as principal component regression (PCR) and partial-least-squares regression (PLSR) may therefore offer a promising alternate approach.[35,36]

The present investigation has a number of purposes, but the primary aim is to model experimentally determined normal boiling points (NBP) of halogenated aliphatic compounds from computationally derived molecular descriptors, and to compare this approach with a group contribution method. A secondary purpose is to investigate if the QSPR model can be used to validate the available experimental database.

DATA AND COMPUTATIONAL METHODS

Experimentally determined values for the normal boiling points of 240 halogenated aliphatic compounds were obtained from the ICI study mentioned above.[19] Experimentally determined boiling points were also obtained for 74 of these 240 compounds from another data source, the PhysProp Database (Syracuse Research Corp., Syracuse, NY).

The chemical structure of each compound was sketched on a PC using the software HyperChem v.7.01 (HyperCube, Inc., Gainesville, FL). Each compound was modeled using the force-field routine MM+, an extension by HyperCube of the standard MM2 force field.[37,38] The molecular structures were then used as input for the generation of 1175 empirical descriptors using the software Dragon v.3.0 (Milano Chemometrics and QSAR Research Group, University of Milano-Bicocca, Milan, Italy). The molecular descriptors thus generated include constitutional descriptors, topological descriptors, molecular walk counts, BCUT descriptors, Galvez topological charge indices, 2D autocorrelations, Randic molecular profiles, geometrical descriptors, RDF descriptors, 3D-MoRSE descriptors, WHIM descriptors, GETAWAY descriptors, functional group counts, and atom-centered fragments. Most of these descriptors are reviewed in the recent textbook by Todeschini and Consonni.[39] Estimations of boiling points using the group contribution method of Stein and Brown with additional extensions were performed using the software MPBPWIN v.1.4 (U.S. Environmental Protection Agency, Washington, DC).

The data analysis and multivariate calibrations were carried out using the software Matlab v.6.5 (Mathworks Inc., Natick, MA), Statistica v.6.1 (StatSoft Inc., Tulsa, OK), and Unscrambler v.7.6 SR-1 (CAMO ASA, Oslo, Norway). Analysis of variance (ANOVA), PCA, PLSR, and implicit nonlinear latent variable regression (INLR) were used as methods for data analysis and modeling. ANOVA is described by Ståhle and Wold.[40] Data analysis with PCA is reviewed by
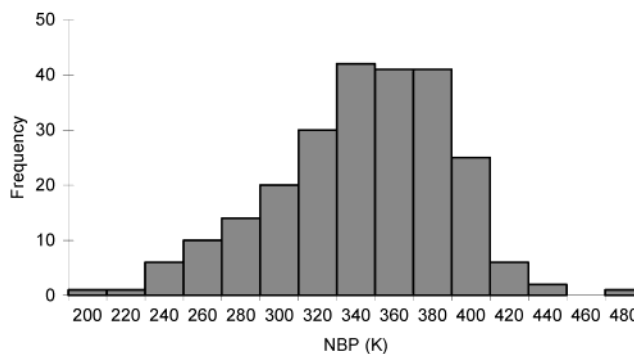

Figure 1. Frequency distribution of experimentally determined normal boiling points for halogenated aliphatic compounds.

Jackson.[41] Multivariate calibration with PLSR is described by Martens and Naes[42] and with INLR by Berglund and Wold.[43] Descriptor variables with minor influence in the PLS regression were assigned zero weight, and these variables were identified using a jackknife method for significance testing of the model parameters.[44]

All descriptor variables were preprocessed by autoscaling to zero mean and unit variance. Cross-validation was used to establish the rank of the calibration model (number of latent variables), and an external test set was used to estimate the prediction error. The calibration model was characterized by the standard deviations of the prediction residuals, for the calibration objects and the external test set respectively: SEC (standard error of calibration) and SEP (standard error of prediction). The explained variances are defined as the sum of squares due to regression divided by the sum of squares about the mean: R^2 (square of the multiple correlation coefficient for the calibration objects) and Q^2 (square of the multiple correlation coefficient for the external test set).

RESULTS AND DISCUSSION

The numbering of compounds used in the report of the ICI study is also applied here.[19] These compounds cover a variety of fluorine, chlorine, bromine, and mixed halogenated substitution patterns (see Table 1).

The boiling points vary between 191 and 462 K; the frequency distribution is shown in Figure 1.

Identification of Outliers. The descriptor data were initially modeled using PCA. A two-component model explained 48% of the variance in the descriptor variables. The first two score vectors are shown in Figure 2.

The PCA scores plot gives a quick overview, and the position of the compounds on this plot is directly related to the chemical structures as described by the molecular descriptors. Two compounds show a substantial deviation from the other 238 halogenated aliphatics: 1,1,1,2,2,3,3,4,4,5,5,6,6,7,7-pentadecafluoroheptane (194) and hexadecafluoroheptane
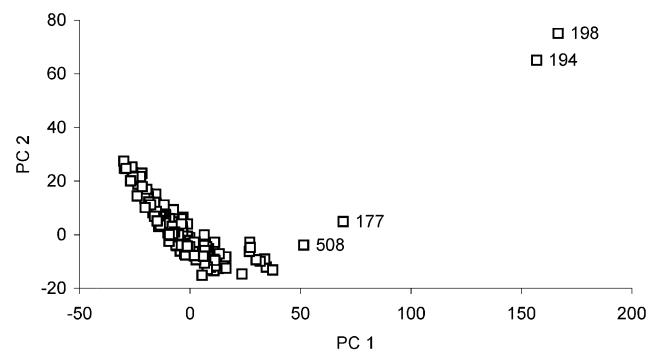
QSPRs OF HALOGENATED ALIPHATIC COMPOUNDS

J. Chem. Inf. Comput. Sci., Vol. 44, No. 1, 2004 **189**

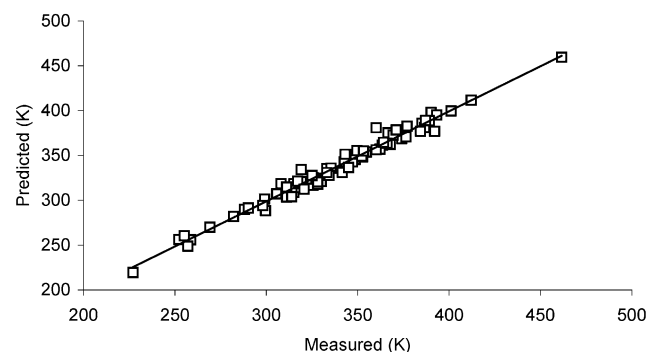**Figure 2.** Scores for the halogenated aliphatic compounds.

Figure 3. Predicted vs measured normal boiling points (K) for the PLSR model. Validation results for 75 halogenated aliphatic compounds in a separate test set.

Figure 4. Explained variance from cross-validation (%), for molecular descriptors and the normal boiling point, vs the number of latent variables in the PLSR model.

Figure 5. Molar volume (cm^3 mol^{-1}) vs first latent variable in the PLSR model.

(198). Since the rest of the data set did not share the same structural features of a long carbon chain and 15−16 fluorine atoms, these outlying objects could obviously not be included into a multivariate calibration model.

Seventeen additional outliers were identified by visual inspection of residual variance and leverage plots from a preliminary PLS regression on all objects. These outliers will be discussed in greater detail below.

Calibration Model. The 221 objects remaining after removal of the 19 outliers were randomly assigned to a calibration set of 146 objects and a test set of 75 objects. Five hundred and eleven descriptor variables were selected for inclusion in the final model based on significance tests using jackknifing in preliminary runs. These descriptors represent all of the previously listed groups, thus supporting the observation by others that there exist no single "best" set of descriptors.[45,46] The number of latent variables to retain in the PLSR model was estimated at six using full (leave-one-out) cross-validation. The six latent variables capture 71.6% of the variance in the 511 descriptor variables, thus demonstrating that the information contained in the descriptors is effectively used in the calibration model.

The model was subsequently validated with the external test set. The model performance parameters were as follows: SEC, 4.90 K; R^2_{cal}, 0.989; SEP, 6.17 K; and Q^2_{ext}, 0.978. Many authors prefer to report the average absolute errors: the calibration error is then 3.72 K and the test set prediction error 4.78 K. Figure 3 shows predicted versus measured results for the 75 objects in the test set.

The standard error of prediction and the average absolute prediction error are both in agreement with the expected lower bound of the experimental error, i.e., 5 K. Many aut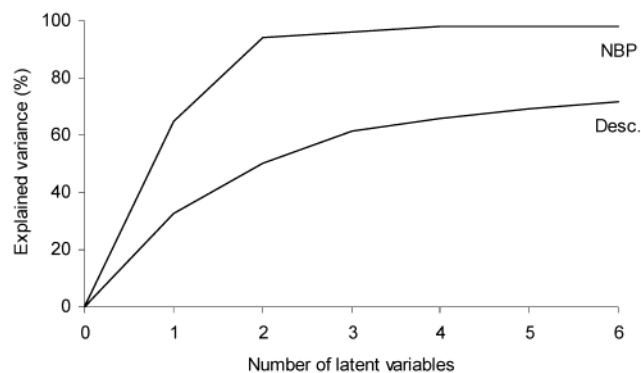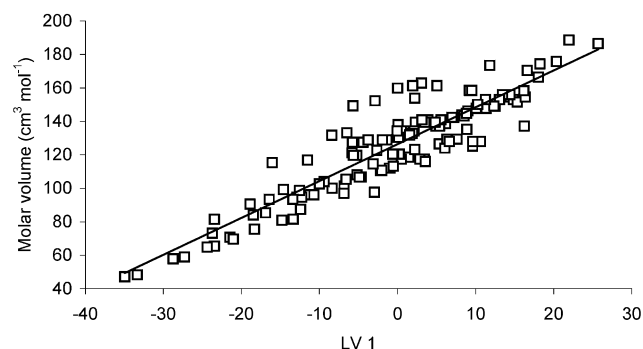hors have suggested that nonlinear techniques such as neural networks are necessary to minimize the prediction error. Recalibration with INLR, an extension of PLSR by expansion of the descriptor matrix with quadratic terms (after mean centering), did not improve the calibration model (SEC, 5.02 K; R^2_{cal}, 0.989). This, and the already low prediction error, suggests that additional parameters or other approaches to describing nonlinearities would not improve the outcome. The bilinear PLSR models can in fact often accommodate nonlinear behavior by the addition of more latent variables.[42]

Main Factors of Variation. Most of the variance both in the molecular descriptors and the normal boiling point is described by the first two latent variables (Figure 4).

The last four latent variables are those that we can assume to account for any nonlinear behavior, which are only 4.3% of the total variance for the boiling point data. A PLSR model with only two latent variables would still perform within the upper bound of the experimental error, i.e., 10 K. A large portion of the variation in the boiling point can be attributed to dispersion forces, and consequently the first latent variable describes mainly variation in size. Figure 5 shows the correlation between the first latent variable and McGowan's molar volume for the calibration data set, calculated by summing the characteristic atomic volumes and correcting for the number of bonds.[47]

The second latent variable can similarly be attributed to the influence of fluorine substitution in decreasing molecular polarizability; Figure 6 shows the correlation with the number of fluorine atoms.

The increase of the boiling point with size and the decrease with fluorine substitution can be described by a simple linear model, accounting for about 70% of the variance. Table 2 shows the results from an analysis of variance (ANOVA)
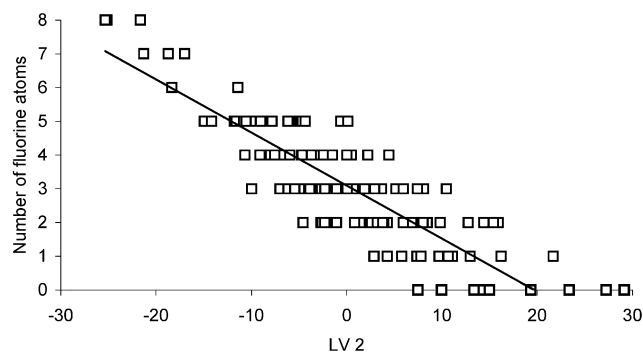
Figure 6. Number of fluorine atoms vs second latent variable in the PLSR model.

Table 2. ANOVA for a Linear Model ($R^2 = 0.689$)

source	SS (×100)	DF	MS (×100)	F	P
model	2222	2	1111	158.3	<0.00001
error	1004	143	7.020		
total	3226	145	22.25		
factors					
molar volume	2154	1	2154	306.9	<0.00001
no. of fluorines	764.9	1	764.9	109.0	<0.00001

for the calibration data set (146 objects). The acronyms used in the ANOVA table are SS for the sum of squares, DF for the degrees of freedom, MS for the mean squares, *F* for the ratio of mean squares, and *P* for the probability in relation to the *F* distribution.

Comparison with a Group Contribution Method. The group contribution method of Stein and Brown was validated for a large number of compounds with an average absolute error for the test set of 20.4 K.[24] The average absolute error of the predictions compared with the experimental data from the ICI study is 23.2 K—thus in close agreement. Predictions with the Stein and Brown method, as implemented in the MPBPWIN software, versus measured results are shown in Figure 7.

Stein and Brown also mentioned estimation problems with perfluorinated compounds.[24] Some fluorocarbons (FCs) and hydrofluorocarbons (HFCs) show here too the largest deviation, i.e., compounds 1,1,1,2,2,3,3-hepta-fluoropropane (76), 1,1,2,2,3,3,4,4-octafluorobutane (164),
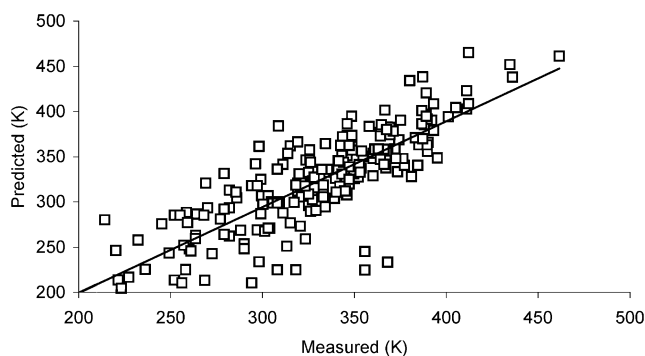

Figure 7. Predicted vs measured normal boiling points (K) using the group contribution method of Stein and Brown with additional extensions. Results for all the 240 halogenated aliphatic compounds: SEP, 30.3 K; and Q^2, 0.664.

1,1,1,2,2,3,3,4,4,5,5,6,6,7,7-pentadecafluoroheptane (194), hexadecafluoroheptane (198), 1,1,1,2,3,4,4,4-octafluorobu-tane (443), and 1,1,2,2,3,3,4,4,5-nonafluoropentane (508).

Validation of Experimental Data. Experimental data reported in the PhysProp Database for 74 compounds were used for comparison. In general the agreement between data from the two sources was excellent, with a deviation between reported values of less than 5 K. Seven objects did, however, deviate more than 10 K as seen from a normal probability plot (Figure 8).

The frequency of deviations greater than 10 K between the two data sources is then 9%, so clearly there is a need for critical evaluation and validation. Nineteen outliers were initially identified in the ICI study, and it is of interest to examine these further. Outliers can be due both to high variance or leverage in the descriptor variables or a high residual variance in the dependent variable. Outliers in the descriptor variables are outside the calibration domain, and model predictions for compounds 194 and 198 are therefore expected to deviate substantially. Experimental data, PLSR model predictions, and group contribution model predictions are all summarized in Table 3.

The results presented in Table 3 show that the PLSR model predictions agree well with the experimental results from the PhysProp database and Carlton, with an average absolute

Table 3. Outliers in PLSR Calibration Modeling: Comparison of Experimentally Determined Boiling Points from the ICI Study and the PhysProp Database with Model Predictions of the PLSR and Group Contribution Methods

no.	compound	ICI (K)	PhysProp (K)	PLSR (K)	group (K)
42	1,1,2-trichloro-1,2-difluoroethane	296.2		346.3	342.2
62	1-chloro-2,2-difluoroethane	326.2	308.3[a]	301.1	300.0
76	1,1,1,2,2,3,3-heptafluoropropane	294.2		255.6	210.5
82	1,1,2,3-tetrachloro-1,3,3-trifluoropropane	358.2		398.9	383.4
126	1,1-dibromo-2,2-difluoropropane	348.6		388.6	394.8
154	2,3-dichloro-hexafluoro-2-butene	381.2	341.7	347.4	328.2
177	dodecafluoropentane	268.7	302.4	296.7	213.3
194	1,1,1,2,2,3,3,4,4,5,5,6,6,7,7-pentadecafluoroheptane	368.2	369.2	385.6	233.4
198	hexadecafluoroheptane	355.7	355.7	363.3	224.8
308	1-bromo-1,1-dichloro-2,2-difluoropropane	308.7		387.1	384.0
314	1,1,1-trichloro-2,2-difluoropropane	326.2	346.2	364.2	343.7
363	1-bromo-2-chloro-1,1,2,2-tetrafluoroethane	347.2		305.6	324.9
364	1,1,1-tribromo-2,2,2-trifluoroethane	387.2		395.6	438.2
367	1,2,2-trichloro-1,1-difluoroethane	389.7	345.1	345.7	355.8
393	1,1-dichloro-1,2,2,3-tetrafluoropropane	346.2		390.6	386.3
408	1,1,1-trichloro-2,2-difluoropropane	326.2	346.2	364.3	343.7
415	1,1,3-tribromo-2,2-difluoropropane	412.2		458.7	465.3
421	1,2-dichloro-2,3-difluoropropane	367.7		398.4	380.0
508	1,1,2,2,3,3,4,4,5-nonafluoropentane	355.7		355.0	245.2

[a] Experimental value from Carlton.[9]

QSPRs OF HALOGENATED ALIPHATIC COMPOUNDS

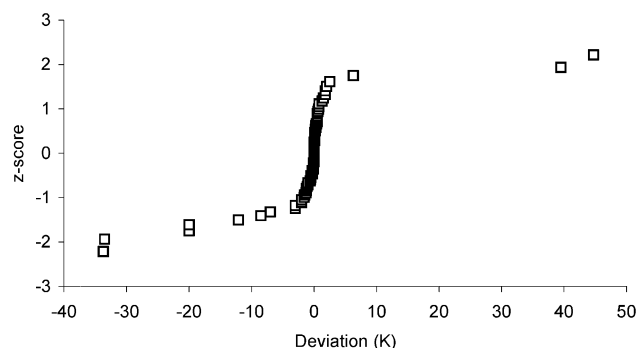*J. Chem. Inf. Comput. Sci., Vol. 44, No. 1, 2004* **191**

Figure 8. Normal probability *z*-scores vs deviations in normal boiling points (K) between the ICI study and the PhysProp database.

error of 9.9 K. The deviation both between the different data sources and with the group contribution method is much larger. It therefore seems advisable to treat the reported experimental results for most of these outliers with caution.

The PLSR model predictions deviate by more than twice the previously reported SEP for 21 compounds. Sixteen of these mismatching predictions belong to the outliers listed above; the other five compounds are 1,2-dibromo-1,1,2,2-tetrafluoroethane (18), 1-bromo-1,1,2,2-tetrafluoroethane (32), 1,3-dibromo-1,1,2,2-tetrafluoropropane (93), 2,2-dichloro-1,3-difluoropropane (122), and 1,1,2-trichloroethene (205). Three of these 21 compounds have measured results reported in the PhysProp database that agree well with the ICI data, thus reducing the mismatch to 18 compounds (nos. 32, 42, 62, 76, 82, 93, 122, 126, 154, 177, 308, 314, 363, 367, 393, 408, 415, and 421). This means that for 92.5% of the ICI data there is good agreement, either with model predictions or with other experimental results. Using the same approach and model with the PhysProp data shows that in this case 96% of the experimental data can be validated using the same standards. The reported boiling points for three of the 74 compounds deviate by more than $2 \times$ SEP and further do not agree with the ICI data: compounds 314 and 408 in Table 3 and 1,2-dibromo-1,1-dichloroethane (50).

CONCLUSIONS

The results from this study show that a bilinear PLSR model is a suitable choice to describe the quantitative structure−property relationship (QSPR) between the normal boiling points and computationally derived molecular descriptors for halogenated aliphatic compounds. The prediction error for a separate test set approaches the anticipated lower bound of experimental error. This is in agreement with the findings of other authors using similar approaches to study other groups of compounds.[20,21,25,29]

Using projection to latent variables instead of traditional methods of variable reduction, such as stepwise regression, facilitates the identification and removal of outliers. The physical−chemical interpretation of the main latent variables indicated, as expected, that the main source of variation can be attributed to molecular size and polarizability.

The QSPR model developed in this study was more precise and accurate than the group contribution method of Stein and Brown. This is also to be expected since this model is optimized for a limited domain, while the group contribution method has a more general applicability. Within this domain the current QSPR model can also be used to predict boiling

points for compounds where they are not available, or even for compounds not yet synthesized. This might prove to be of value, both for technical development work and also because of the environmental relevance of these compounds.

QSPR models can be used to check and validate experimental data.[48,49] The results presented in this study indicate that this is also a viable approach here. Comparison with experimental data from another source seems to support the accuracy of the model predictions. This modeling approach can readily be extended to other physical properties as soon as sufficient calibration data are available.

REFERENCES AND NOTES

(1) Kinney, C. R. A system correlating molecular structure of organic compounds with their boiling points. I. Aliphatic boiling point numbers. *J. Am. Chem. Soc.* **1938**, *60*, 3032−3039.

(2) Wessel, M. D.; Jurs, P. C. Prediction of normal boiling points for a diverse set of industrially important organic compounds from molecular structure. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 841−850.

(3) Molina, M. J.; Rowland, F. S. Stratospheric sink for chlorofluoromethanes−chlorine atomic catalyzed destruction of ozone. *Nature* **1974**, *249*, 810−812.

(4) Manahan, S. E. *Environmental chemistry*; 7th ed.; Lewis Publishers: Boca Raton, FL, 1999.

(5) *Handbook for the international treaties for the protection of the ozone layer*, 5th ed.; United Nations Environment Programme: Geneva, Switzerland, 2000.

(6) *A guide to the climate change convention and its Kyoto protocol*; The Secretariat of the United Nations Framework Convention on Climate Change: Bonn, Germany, 2002.

(7) Woolf, A. A. Predicting boiling points of hydrofluorocarbons. *J. Fluorine Chem.* **1996**, *78*, 151−154.

(8) Yoshida, H.; Miyashita, Y.; Sasaki, S. Modelling of halomethanes using neural networks. *Chemometr. Intell. Lab.* **1996**, *32*, 193−199.

(9) Carlton, T. S. Correlation of boiling points with molecular structure for chlorofluoroethanes. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 158−164.

(10) Balaban, A. T.; Joshi, N.; Kier, L. B.; Hall, L. H. Correlations between chemical structure and normal boiling points of halogenated alkanes C-1−C-4. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 233−237.

(11) Le, T. D.; Weers, J. G. Group contribution-additivity and quantum-mechanical models for predicting the molar refractions, indexes of refraction, and boiling points of fluorochemicals. *J. Phys. Chem.* **1995**, *99*, 13909−13916.

(12) Balaban, A. T.; Basak, S. C.; Colburn, T.; Grunwald, G. D. Correlation between structure and normal boiling points of haloalkanes C-1−C-4 using neural networks. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1118−1121.

(13) Le, T. D.; Weers, J. G. QSPR and GCA models for predicting the normal boiling points of fluorocarbons. *J. Phys. Chem.* **1995**, *99*, 6739−6747.

(14) Woolf, A. A. Relative boiling points of fluoroethers, fluoroamines, and other fluorocarbon derivatives to fluorocarbons. *J. Fluorine Chem.* **1999**, *94*, 47−50.

(15) Wei, J. Boiling points and melting points of chlorofluorocarbons. *Ind. Eng. Chem. Res.* **2000**, *39*, 3116−3119.

(16) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. Normal boiling points for organic compounds: Correlation and prediction by a quantitative structure−property relationship. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 28−41.

(17) Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. Quantitative structure−property relationship study of normal boiling points for halogen/oxygen/sulfur-containing organic compounds using the CODESSA program. *Tetrahedron* **1998**, *54*, 9129−9142.

(18) Gakh, A. A.; Gakh, E. G.; Sumpter, B. G.; Noid, D. W.; Trowbridge, L. D.; Harkins, D. A. Estimation of the properties of hydrofluorocarbons by computer neural networks. *J. Fluorine Chem.* **1995**, *73*, 107−111.

(19) Horvath, A. L. Boiling points of halogenated organic compounds. *Chemosphere* **2001**, *44*, 897−905.

(20) Egolf, L. M.; Wessel, M. D.; Jurs, P. C. Prediction of boiling points and critical temperatures of industrially important organic compounds from molecular structure. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 947−956.

(21) Wessel, M. D.; Jurs, P. C. Prediction of normal boiling points of hydrocarbons from molecular structure. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 68−76.

192 *J. Chem. Inf. Comput. Sci., Vol. 44, No. 1, 2004*

ÖBERG

(22) Horvath, A. L. *Molecular design: Chemical structure generation from the properties of pure organic compounds*; Elsevier: Amsterdam, 1992.

(23) Joback, K. G.; Reid, R. C. Estimation of pure-component properties from group-contributions. *Chem. Eng. Commun.* **1987**, *57*, 233−243.

(24) Stein, S. E.; Brown, R. L. Estimation of normal boiling points from group contributions. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 581−587.

(25) Stanton, D. T. Development of a quantitative structure−property relationship model for estimating normal boiling points of small multifunctional organic molecules. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 81−90.

(26) Katritzky, A. R.; Karelson, M.; Lobanov, V. S. QSPR as a means of predicting and understanding chemical and physical properties in terms of structure. *Pure Appl. Chem.* **1997**, *69*, 245−248.

(27) Karelson, M. *Molecular descriptors in QSAR/QSPR*; Wiley-Interscience: New York, 2000.

(28) Seasholtz, M. B.; Kowalski, B. The parsimony principle applied to multivariate calibration. *Anal. Chim. Acta* **1993**, *277*, 165−177.

(29) Goll, E. S.; Jurs, P. C. Prediction of the normal boiling points of organic compounds from molecular structures with a computational neural network model. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 974−983.

(30) Chalk, A. J.; Beck, B.; Clark, T. A quantum mechanical/neural net model for boiling points with error estimation. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 457−462.

(31) Goll, E. S.; Jurs, P. C. Prediction of vapor pressures of hydrocarbons and halohydrocarbons from molecular structure with a computational neural network model. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1081−1089.

(32) Espinosa, G.; Yaffe, D.; Cohen, Y.; Arenas, A.; Giralt, F. Neural network based quantitative structural property relations (QSPRs) for predicting boiling points of aliphatic hydrocarbons. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 859−879.

(33) Lucic, B.; Trinajstic, N. A new efficient approach for variable selection based on multiregression: Prediction of gas chromatographic retention times and response factors. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 610−621.

(34) Lucic, B.; Amic, D.; Trinajstic, N. Nonlinear multivariate regression outperforms several concisely designed neural networks on three QSPR data sets. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 403−413.

(35) Katritzky, A. R.; Petrukhin, R.; Tatham, D.; Basak, S.; Benfenati, E.; Karelson, M.; Maran, U. Interpretation of quantitative structure−property and −activity relationships. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 679−685.

(36) Öberg, T. Prediction of vapour pressures for halogenated diphenyl ether congeners from molecular descriptors. *Environ. Sci. Pollut. Res.* **2002**, *9*, 405−411.

(37) Allinger, N. L. Conformational analysis 130. MM2−Hydrocarbon force-field utilizing V1 and V2 torsional terms. *J. Am. Chem. Soc.* **1977**, *99*, 8127−8134.

(38) Hocquet, A.; Langgard, M. An evaluation of the MM+ force field. *J. Mol. Model.* **1998**, *4*, 94−112.

(39) Todeschini, R.; Consonni, V. *Handbook of molecular descriptors*; Wiley-VCH: Weinheim, Germany, 2000.

(40) Ståhle, L.; Wold, S. Analysis of variance (ANOVA). *Chemometr. Intell. Lab.* **1990**, *6*, 259−272.

(41) Jackson, J. E. *A user's guide to principal components*; Wiley: New York, 1991.

(42) Martens, H.; Næs, T. *Multivariate calibration*; Wiley: Chichester, U.K., 1989.

(43) Berglund, A.; Wold, S. INLR, implicit non-linear latent variable regression. *J. Chemometr.* **1997**, *11*, 141−156.

(44) Martens, H.; Martens, M. Modified jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food. Qual. Prefer.* **2000**, *11*, 5−16.

(45) Livingstone, D. J. The characterization of chemical structures using molecular properties. A survey. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 195−209.

(46) Consonni, V.; Todeschini, R.; Pavan, M. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 682−692.

(47) Abraham, M. H.; McGowan, J. C. The use of characteristic volumes to measure cavity terms in reversed phase liquid chromatography. *Chromatographia* **1987**, *23*, 243−246.

(48) Schwarzenbach, R. P.; Gschwend, P. M.; Imboden, D. M. *Environmental organic chemistry*, 2nd ed.; Wiley: Hoboken, NJ, 2002.

(49) Öberg, T. Prediction of physical properties for PCB congeners from molecular descriptors. *Internet J. Chem.* **2001**, *4*, 11.