

QSPR Using MOLGEN-QSPR: The Challenge of Fluoroalkane Boiling Points

Christoph Rücker,* Markus Meringer, Adalbert Kerber
Department of Mathematics, Universität Bayreuth,
D-95440 Bayreuth, Germany

Received ...

By means of the new software MOLGEN-QSPR, a multilinear regression model for the boiling points of fluoroalkanes was established. The model is based exclusively on simple descriptors derived directly from molecular structure, and nevertheless describes a broader set of data more precisely than previous attempts that used either more demanding (quantum chemical) descriptors or more demanding (nonlinear) statistical methods such as neural networks. The model's internal consistency was confirmed by leave-one-out crossvalidation. The model was used to predict all unknown boiling points of fluorobutanes, and the quality of predictions was estimated by means of comparison with boiling point predictions for fluoropentanes.

INTRODUCTION

Recent innovative use of poly- or perfluorinated alkanes has fostered interest in the physical properties of fluoroalkanes.¹ A compound's normal boiling point (bp) is a fundamental thermodynamic property, interesting both in its own right and as a basis for the calculation of other properties.² In previous studies it was noticed that fluoroalkanes exhibit boiling points that are more difficult to describe or predict in terms of molecular structure than those of alkanes or other haloalkanes.³⁻⁷ This follows from the particular properties of the fluorine atom compared to the other halogens, F being both highly electronegative and a rather small and "hard" sphere of low atomic weight. A C-F bond is a strong dipole, and these local dipoles interact intra- and intermolecularly even if the resulting dipole moment of an isolated molecule is zero. A consequence is the decidedly nonlinear dependence of fluoroalkane boiling points on the number of fluorine atoms, resulting in a bp maximum about halfway between an alkane and the corresponding perfluoroalkane (CH_4 -161.5°C, CH_3F -78.5, CH_2F_2 -51.6, CHF_3 -82, CF_4 -128).^{3a}

For this reason previous researchers used inherently nonlinear methods such as neural networks for the description of fluoroalkane boiling points.^{3b,4} A manual graphical fitting method was used to describe the bps of fluoroethanes or, separately, of fluoropropanes, and allowed some predictions.⁵ Portability, parsimony, and predictive power of these methods, however, are less than desirable. Alternatively, multilinear regression (MLR) models required six mostly electrostatic and quantum chemical descriptors for describing the bps of 42 fluoroalkanes ($s = 6.3^\circ\text{C}$).⁸ Accepting fluoroalkane boiling points as a challenge we were curious on the viability of a model of the simplest kind for the boiling points of this class of compounds, i.e. a multilinear regression model based on simple descriptors that are directly derived from molecular structure. This is by no means an easy task, as illustrated by the fact that for fluorobutanes alone boiling points between -2 and $+110^\circ\text{C}$ have been reported, and even among the restricted set of tetrafluorobutanes bps vary between 17 and 110°C .

METHODS

Data. Boiling points at normal pressure of $\text{C}_1 - \text{C}_4$ fluoroalkanes (herein given in $^\circ\text{C}$) were taken from the literature^{3,5,9-11} and checked against the Beilstein database in order to avoid fitting wrong data. A few compounds/boiling points found in Beilstein only were also included. In cases of marginal divergence between boiling points reported in the sources the average was taken. In cases of major divergence boiling points were excluded. This resulted in a data set of 83 $\text{C}_1\text{-C}_4$ fluoroalkanes with boiling points, i.e. all 42 possible fluoromethanes through fluoropropanes plus 41 (out of 116 possible, ignoring stereoisomerism) fluorobutanes. We did not normally examine the primary sources given in Beilstein, and thus cannot exclude the remote possibility that a few bp values accepted here may be calculated rather than experimental values.¹² This is the most comprehensive collection of lower fluoroalkane bps of which we are aware.

Boiling points were attributed reliability classes as in our earlier studies:^{7,13} Boiling points appearing in the Beilstein database only once are in reliability class 0, those measured at least twice by independent researchers with a difference of at most 4°C are in class 1, those measured at least four times by independent authors and differing no more than 2°C are in reliability class 2. According to this measure, the data

used in the present study are on average of lower quality than those in our previous studies, with many fluoroalkane bps reported only once. Information on reliability, though not used in the calculations, proved useful for identification of dubious bp data and in the selection of reference data.

The bp reported for 1,2-difluorobutane, $\text{FH}_2\text{C}-\text{CHF}-\text{CH}_2-\text{CH}_3$, 39°C , reliability class 0, is obviously unreasonable and was therefore excluded.¹⁴ For 1,2-difluoroethane, $\text{FH}_2\text{C}-\text{CH}_2\text{F}$, we used the bp value 10.5°C , as was done in most previous studies.^{4-6,15}

Descriptor Calculation. We used our software MOLGEN-QSPR that combines structure generation with calculation of many molecular descriptor values and with data treatment by various statistical methods.^{7,12} For a set of compounds structurally rather homogenous such as the fluorinated alkanes, substructure and fragment counts could be anticipated to be valuable descriptors.¹⁶ In fact, in previous studies such descriptors had been extensively used.^{5,6} So, in addition to the descriptors offered routinely by MOLGEN-QSPR,¹⁷ we included in this study substructure counts and fragment counts.¹⁸

Substructure counts are automatically provided by MOLGEN-QSPR, a substructure is a part of the hydrogen-suppressed molecular graph. For example, counts of all substructures of one to four bonds were requested, and these substructures were generated and counted: C-C, C-F, C-C-C, F-C-F, C-C-F, C-C-C-C, C-C(C)-C, $\text{F}_2\text{C}-\text{C}$, F-C-C-F, F_3C , C-C-C-F, C-C(C)-F, $\text{F}_3\text{C}-\text{C}$, $\text{F}_2\text{C}-\text{C}-\text{F}$, C-C(F)-C-C, C-C-C-C-F, C-C(C)(C)-F, C-C(C)-C-F, F-C-C-C-F, F-C-C(C)-F, $\text{F}_2\text{C}-\text{C}-\text{C}$, C-C(C)(F)-F. MOLGEN rule for substructure counts: Two realizations of the same substructure are counted separately if they differ in at least one bond, i.e. if they differ in at least one (non-H) atom.

In contrast, a fragment is a part of the full molecular graph, it may therefore include hydrogen atoms. A fragment is defined by the user and searched for by the system on request. For example, counts of the following fragments were used in this study:

H-C-F, H-C-C-F (according to Woolf^{5b});

CHF_2 , CH_2F (according to Carlton⁶);

CH_3 , C- CH_2 -C, C-CH(C)-C, C-CHF-C, $\text{FH}_2\text{C}-\text{CHF}$, FHC-CHF.

MOLGEN rule for fragment counts: Two realizations of the same fragment are counted separately if they differ in at least one non-H atom.

For example, by this rule 1,1,1,2,3-pentafluoropropane, $F_3C-CHF-CH_2F$, contains 2 H-C-F and 5 H-C-C-F fragments.

In Woolf's study^{5b} realizations of a fragment are counted separately if they differ in at least one atom (which may be an H atom). By this rule $F_3C-CHF-CH_2F$ contains 3 H-C-F fragments and 6 H-C-C-F fragments. Counts of H-C-F and H-C-C-F fragments obtained by this rule (HCF_{man} and $HCCF_{man}$) were manually added to our data.

Further descriptors based on physical concepts were included in the descriptor pool:

$bip = n(CHF_2) + n(CH_2F)$ (bipolar groups, Carlton⁶);

$tbip = bip + n(C-CHF-C)$ (total bipolar groups);

$xsF = N_F - n(CH_2F) - N_H + n(CHF_2)$, unless this number is negative, in which case xsF is set to 0 (excess exterior fluorine atoms⁶);

$n(CF_3)^2$, $n(CH_3)^2$, $n(CF_3) \cdot n(CH_3)$ (Woolf^{5b}).

Finally, descriptors $(relN_F)^2 = (N_F/\text{number of all atoms})^2$ and $(Frate)^2 = (N_F/(2N_C+2))^2$ were included to account for the bp maximum mentioned in the introduction.

In the step-up procedure⁷ the parameter N_{model} was always set to 1000.

RESULTS AND DISCUSSION

Preliminary study. For comparison with the results of Ivanciuc et al.,⁸ we initially treated their data using our descriptors and statistical procedures. For this purpose compound structures and boiling points were taken from Balaban et al.,^{3a} as described in ref. 8.

Unfortunately, Ivanciuc et al. reported results for a set of 42 C_1-C_4 fluoroalkanes, whereas Balaban's data^{3a} include 43 fluoroalkanes. So one compound is missing in Ivanciuc's study, but we do not know which one.

For the 43 fluoroalkanes MOLGEN-QSPR found the best 6-descriptor MLR model as follows. (In the text we characterize a MLR model by the descriptors involved and by its r^2 , s , F , r^2_{cv} and s_{cv} values, the latter two refer to leave-one-out crossvalidation. For full models see Table 1.)

Xu^m , ${}^0\chi^v$, ${}^4\chi_p$, $S(ssssC)$, $n(F-C-F)$, $n(F-C-C-F)$ (model 0a)

$r^2 = 0.9883$, $s = 4.624$, $F = 508$, $r^2_{cv} = 0.9826$, $s_{cv} = 5.652$, $N = 43$.^{19,20}

Xu^m is the modified Xu index,^{17,21} $S(ssssC)$ is the sum of electrotopological state indices of carbon atoms bearing no hydrogen,²² $n(F-C-F)$ and $n(F-C-C-F)$ are counts of the respective substructures.¹⁸

This result is to be compared to the previously best 6-descriptor MLR model, which comprises one topological, two electrostatic, and three quantum-chemical descriptors:

$$r = 0.989 \ (r^2 = 0.9781), \ s = 6.3, \ F = 267, \ N = 42.^8$$

So even without any quantum-chemical descriptor better results than in reference 8 are achievable.

Main study. For the larger set of 82 C₁-C₄ fluoroalkane bps ([83fluoroalkanes.md4](#)) the step-up procedure found the best multilinear 6-descriptor model

$$Xu^m, \ ^2TC^v, \ ^3TM1, \ HCCF_{man}, \ (relN_F)^2, \ n(FH_2C-CHF)=vicdiF \quad (\text{model 1})$$

$$r^2 = 0.9845, \ s = 5.140, \ F = 793, \ r^2_{cv} = 0.9812, \ s_{cv} = 5.679, \ N = 82.$$

([83fluoroalkanesmodell1.qspr](#))

²TC^v and ³TM1 are Bonchev overall indices,²³ HCCF_{man} is the number of HCCF fragments as counted by Woolf,^{5b} and n(FH₂C-CHF) is the count of fragment FH₂C-CHF.

The best 7-descriptor model found is

$$Xu^m, \ ^2P, \ S(ssssC), \ n(F-C-F)=sub4, \ n(F-C-C-F)=sub9, \ (relN_F)^2, \ n(FH_2C-CHF)=vicdiF \quad (\text{model 2})$$

$$r^2 = 0.9872, \ s = 4.701, \ F = 815, \ r^2_{cv} = 0.9840, \ s_{cv} = 5.252, \ N = 82.$$

([83fluoroalkanesmodel2.qspr](#))

²P is the number of paths of length 2 in the H-suppressed molecular graph.

Figure 1 is a plot of experimental vs calculated bps (by model 2) and of the corresponding bps obtained by leave-one-out crossvalidation.

(Figure 1)

Table 2 ([83fluoroalkanesTab.txt](#)) lists experimental and calculated bps (by model 2) and residuals.

(Table 2)

Figure 2 visualizes the results separately for fluorinated methanes, ethanes, propanes, and butanes. By inspection, the quality of data description is similar in all four subpopulations. In fact, the average absolute error for the 4 fluoromethanes, 9 fluoroethanes, 29

fluoropropanes, and 40 fluorobutanes is 3.72, 3.13, 4.02, and 3.40°C, respectively.

(Figure 2)

Note that models 1 and 2 use both (electro)topological indices and substructure and fragment counts. In fact, (electro)topological indices alone or substructure and fragment counts alone led to best 6-descriptor models of $s = 5.83$ or 6.09 , respectively. As in our earlier study,⁷ geometrical indices did not qualify to appear in the best models. Remarkably also most of the descriptors based on physical concepts^{5b,6} did not qualify for the best models.

Prediction. By means of model 2 we predicted the normal boiling points of those 76 fluorobutanes whose bps are either not at all known or were excluded from our data set for contradicting reports in Beilstein. Table 3 ([Predictions76Fbutanes.txt](#)) shows the predictions.

(Table 3)

Of course, the quality of these predictions cannot be assessed at present. We expect the average absolute error to be a bit larger than, but similar to that of the fit for fluorobutanes above, 3.40°C. An independent estimation was obtained as follows: The data for model 2 lead one to suspect that bp predictions for fluoropentanes by model 2 might not be senseless. Since these are extrapolations, their average error should be larger than that of the predictions for fluorobutanes, which are interpolations. So we predicted by model 2 the bps of all those fluoropentanes that have reliably known experimental bps (reliability classes 1 or 2, eighteen compounds). Results are shown in Table 4 ([18Fpentanes.txt](#)). The average absolute error of these predictions is 4.91°C. The predictions for the fluorobutanes therefore should on average be better than that.

(Table 4)

CONCLUSION

The combination of global descriptors such as topological indices on the one hand and substructure and fragment counts on the other has proven itself to be very useful in QSPR work.^{12,16} Fragment counts offer the scientist some flexibility to react whenever in a QSPR equation several compounds containing a particular structural element have similar residuals. In fact, in this study fragment count $n(\text{FH}_2\text{C}-\text{CHF})$ was included in the descriptor pool after preliminary QSPR models had resulted in large positive residuals for compounds $\text{FH}_2\text{C}-\text{CHF}-\text{CH}_3$, $\text{FH}_2\text{C}-\text{CHF}-\text{CH}_2\text{F}$, $\text{F}_2\text{HC}-\text{CHF}-\text{CH}_2\text{F}$, $\text{FH}_2\text{C}-\text{CHF}-\text{CHF}-\text{CH}_2\text{F}$.

REFERENCES AND NOTES

- (1) Tzschucke, C. C.; Markert, C.; Bannwarth, W.; Roller, S.; Hebel, A.; Haag, R. Modern Separation Techniques for Efficient Workup in Organic Synthesis. *Angew. Chem. Int. Ed.* **2002**, *41*, 3964-4000.
- (2) (a) Reid, C. R.; Prausnitz, J. M.; Sherwood, J. K. The Properties of Gases and Liquids, 4th edition, McGraw-Hill, New York, 1987. (b) Horvath, A. L. *Molecular Design*, Elsevier, Amsterdam, 1992.
- (3) (a) Balaban, A.T.; Joshi, N.; Kier, L.B.; Hall, L.H. Correlations between Chemical Structure and Normal Boiling Points of Halogenated Alkanes C₁ - C₄. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 233-237. (b) Balaban, A. T.; Basak, S. C.; Colburn, T.; Grunwald, G. D. Correlation between Structure and Normal Boiling Points of Haloalkanes C₁-C₄ Using Neural Networks. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1118-1121.
- (4) Gakh, A. A.; Gakh, E. G.; Sumpster, 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.
- (5) (a) Woolf, A. A. Boiling Point Relations in the Halogenated Ethane Series, *J. Fluorine Chem.* **1990**, *50*, 89-99. (b) Woolf, A. A. Predicting Boiling Points of Hydrofluorocarbons, *J. Fluorine Chem.* **1996**, *78*, 151-154.
- (6) Carlton, T. S. Correlation of Boiling Points with Molecular Structure for Chlorofluoroethanes. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 158-164.
- (7) Rücker, C.; Meringer, M.; Kerber, A. QSPR Using MOLGEN-QSPR: The Example of Haloalkane Boiling Points, submitted to *J. Chem. Inf. Comput. Sci.*
- (8) 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.
- (9) Horvath, A. L. Boiling Points of Halogenated Organic Compounds. *Chemosphere*, **2001**, *44*, 897-905.
- (10) The NIST Chemistry Webbook, <http://webbook.nist.gov>.
- (11) Burdon, J.; Garnier, L.; Powell, R.L. Fluorination of Propane over Cobalt(III) trifluoride and Potassium tetrafluorocobaltate(III). *J. Chem. Soc., Perkin Trans. 2*, **1996**, 625-631.
- (12) Kerber, A.; Laue, R.; Meringer, M.; Rücker, C. MOLGEN-QSPR, A Software Package for the Study of Quantitative Structure-Property Relationships. *MATCH Commun. Math. Comput. Chem.* **2004**, *51*, 187-204.

(13) Rücker, G.; Rücker, C. On Topological Indices, Boiling Points, and Cycloalkanes. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 788-802.

(14) By obviously unreasonable we understand, e.g., a bp value that violates the fundamental rule of a considerable increase in bp on enlarging a molecule by a CH₂ group. Consider the following sequences (all bps in reliability classes 1 or 2):

CH₃-CH₂F -37.5, CH₃-CHF-CH₃ -9.7, CH₃-CHF-CH₂-CH₃ 25;

CH₃-CHF₂ -25, CH₃-CF₂-CH₃ -0.5, CH₃-CF₂-CH₂-CH₃ 31;

FH₂C-CH₃ -37.5, FH₂C-CH₂-CH₃ -2.5, FH₂C-CH₂-CH₂-CH₃ 32.5.

From these values the bp increment for a CH₂ group can be estimated to be 25-35°C (for additional examples see reference 14a). Now consider FH₂C-CH₂F, bp 10.45 or 27 (reliability class unattributable¹⁵), FH₂C-CHF-CH₃ 32 (reliability class 0), FH₂C-CHF-CH₂-CH₃ 39 (reliability class 0). There is logic in this sequence only if for FH₂C-CH₂F 10.45, and for FH₂C-CHF-CH₃ 32 are essentially correct. For FH₂C-CHF-CH₂-CH₃ 39°C^{14b} then is far too low, this compound/bp therefore was excluded. (a) Balaban, A. T.; Basak, S. C.; Mills, D. Normal Boiling Points of 1,ω-Alkanedinitriles: The highest Increment in a Homologous Series. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 769. (b) Baklouti, A.; El Gharbi, R. *J. Fluorine Chem.* **1979**, *13*, 297-314.

(15) The bp of FH₂C-CH₂F is somewhat problematic: An experimental bp of 10-11°C was reported only once,^{15a} while there are several reports of bp values between 25 and 31°C for this compound, though not all of these are independent.^{15b-h} However, the former number fits much better into various bp models. In the NIST Chemistry Webbook, 10.45°C is the only bp value given for FH₂C-CH₂F.¹⁰

(a) 10-11, Henne, A. L.; Renoll, M. W. *J. Am. Chem. Soc.* **1936**, *58*, 889.

(b) 30.7, Edgell, W. F.; Parts, L. *J. Am. Chem. Soc.* **1955**, *77*, 4899.

(c) 26-26.2, Titow, *Dokl. Akad. Nauk SSSR* **1957**, *113*, 358.

(d) 31, Klabe, P.; Nielsen, J. R. *J. Chem. Phys.* **1960**, *33*, 1764-1774.

(e) 26, Schiemann, G.; Cornils, B. *Chem. Ber.* **1965**, *98*, 3418-3435.

(f) 26.5, Abraham, R. J.; Kemp, R. H. *J. Chem. Soc. B* **1971**, 1240-1245.

(g) 25-27, Middleton, W. J. *J. Org. Chem.* **1975**, *40*, 574-578.

(h) 30.7, Nappa, N. J.; Sievert, A. C., *J. Fluorine Chem.* **1993**, *62*, 111-118.

(16) Zefirov, N. S.; Palyulin, V. A. Fragmental Approach in QSPR. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1112-1122.

(17) Rücker, C.; Braun, J.; Kerber, A; Laue, R. The Molecular Descriptors Computed with MOLGEN. <http://www.mathe2.uni-bayreuth.de/molgenspr>.

(18) (a) Rücker, G.; Rücker, C. Automatic Enumeration of All Connected Subgraphs. *MATCH Commun. Math. Comput. Chem.* **2000**, *41*, 145-149. (b)

Rücker, G.; Rücker, C. On Finding Nonisomorphic Connected Subgraphs and Distinct Molecular Substructures. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 314-320.

(19) The bp used in the preliminary study for 1,2-difluoroethane is 26°C (as, presumably, in reference 8), and this leads to the by far highest residual, 16.2°C. If 1,2-difluoroethane is excluded, for the remaining 42 compounds the best 6-descriptor model found is

$Xu, {}^0\chi^v, {}^4\chi_p, S(ssssC), n(F-C-F), n(F-C-C-F)$ (model 0b)

$r^2 = 0.9931, s = 3.580, F = 839, r_{cv}^2 = 0.9897, s_{cv} = 4.384, N = 42.$

When instead fluoromethane was excluded (for this compound the flexibility index Φ is undefined, which may have been a reason for exclusion), then the best model we found is

$Xu, {}^0\chi^v, {}^4\chi_p, S(ssssC), n(F-C-F), n(F-C-C-F)$ (model 0c)

$r^2 = 0.9874, s = 4.667, F = 456, r_{cv}^2 = 0.9807, s_{cv} = 5.771, N = 42.$

(20) In the preliminary study fragment counts and the additional descriptors such as bip and xsF were not included in the descriptor pool. If they are included, then models of even slightly better r^2 and s values can be found, which however exhibit worse r_{cv}^2 and s_{cv} values.

(21) Ren, B. A New Topological Index for QSPR of Alkanes. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 139-143.

(22) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*. Academic Press, San Diego, 1999.

(23)(a) Bonchev, D.; Trinajstic, N. Overall Molecular Descriptors. 3. Overall Zagreb Indices. *SAR QSAR Environ. Res.* **2001**, *12*, 213-236. (b)

Bonchev, D. Overall Connectivity - A next Generation Molecular Connectivity. *J. Mol. Graphics Modell.* **2001**, *20*, 65-75.

Footnote for page 1

*Corresponding author phone: +49 921 553386; fax: +049 921 553385; e-mail: christoph.ruecker@uni-bayreuth.de

Figure and Table Captions

Figure 1. Plot of calculated (by model 2, white disks, and by leave-one-out crossvalidation, black disks) vs experimental boiling points of 82 fluoromethanes through fluorobutanes.

Figure 2. Plot of calculated (by model 2, white disks) and experimental (black disks) boiling points of lower fluoroalkanes vs number of F atoms in the molecule.

Table 1. Full MLR models for the boiling points of lower fluoroalkanes.

Table 2. Experimental and calculated boiling points (by model 2), residuals and structures of compounds in the learning set.

Table 3. Predicted boiling points (by model 2) for 76 fluorobutanes not included in the learning set.

Table 4. Experimental (with reliability) and predicted boiling points (extrapolations by model 2), residuals and structures of 18 fluoropentanes.

Table 1. Full MLR models for the boiling points of fluoroalkanes C₁ - C₄.

$$\text{bp} = 83.2226 \cdot \text{Xu}^{\text{m}} - 25.1841 \cdot {}^0\chi^{\text{v}} - 12.2045 \cdot {}^4\chi_{\text{p}} - 5.21304 \cdot \text{S}(\text{ssssC}) \\ - 34.247 \cdot \text{n}(\text{F-C-F}) - 9.2969 \cdot \text{n}(\text{F-C-C-F}) - 41.3515 \quad (\text{model 0a})$$

$$\text{bp} = 67.3463 \cdot \text{Xu}^{\text{m}} - 1.13857 \cdot {}^2\text{TC}^{\text{v}} + 0.092004 \cdot {}^3\text{TM1} + 2.35049 \cdot \text{HCCF}_{\text{man}} \\ - 81.9815 \cdot (\text{relN}_{\text{F}})^2 + 8.5278 \cdot \text{n}(\text{FH}_2\text{C-CHF}) - 75.3646 \quad (\text{model 1})$$

$$\text{bp} = 64.6756 \cdot \text{Xu}^{\text{m}} - 5.51401 \cdot {}^2\text{P} - 2.05454 \cdot \text{S}(\text{ssssC}) - 18.0512 \cdot \text{n}(\text{F-C-F}) \\ - 3.88042 \cdot \text{n}(\text{F-C-C-F}) - 53.7732 \cdot (\text{relN}_{\text{F}})^2 + 7.49879 \cdot \text{n}(\text{FH}_2\text{C-CHF}) - 72.4514 \\ (\text{model 2})$$

Table 2. Experimental and calculated boiling points (by model 2), residuals and structures of compounds in the learning set.

	bp	calc	residual	structure
1	-15	-18.461	3.4608	F3C-CF2-CH3
2	12	8.8309	3.1691	F3C-CH(CH3)2
3	24.5	34.612	-10.112	F3C-CH2-CH2-CF3
4	32.5	33.345	-0.84489	F2HC-CHF-CF2-CF3
5	35	36.325	-1.3245	F2HC-CF2-CHF-CF3
6	26.5	28.468	-1.9683	FH2C-CF2-CF2-CF3
7	-51.6	-57.615	6.015	CH2F2
8	41	38.948	2.0525	FH2C-CH2-CH2F
9	21	24.266	-3.2656	F3C-CHF-CH2F
10	-78.5	-74.602	-3.8977	CH3F
11	11	12.337	-1.3371	F2HC-CF2-CHF2
12	7.5	6.8761	0.62385	F2HC-CH2-CH3
13	6	8.5232	-2.5232	F3C-CHF-CHF2
14	-9.7	-7.9039	-1.7961	H3C-CHF-CH3
15	-13	-11.543	-1.4566	F3C-CH2-CH3
16	-19	-15.058	-3.9422	F3C-CHF-CF3
17	-78	-80.728	2.7278	F3C-CF3
18	-82	-80.509	-1.4914	CHF3
19	-128	-124.53	-3.4678	CF4
20	-2.5	5.2042	-7.7042	FH2C-CH2-CH3
21	-38	-40.317	2.3172	F3C-CF2-CF3
22	-0.5	-8.2093	7.7093	H3C-CF2-CH3
23	29.4	19.518	9.8823	FH2C-CH2-CF3
24	-0.8	-1.8548	1.0548	F3C-CH2-CF3
25	-17	-14.227	-2.7734	F3C-CF2-CHF2
26	4	8.7202	-4.7202	FH2C-CHF2
27	32.5	40.107	-7.6066	FH2C-CH2-CH2-CH3
28	25	26.239	-1.2392	H3C-CH2-CHF-CH3
29	18	20.385	-2.3849	F3C-CH2-CF2-CF3
30	44	40.696	3.3041	FH2C-CF2-CF2-CHF2
31	15	16.08	-1.0797	F2HC-CF2-CF2-CF3
32	-1.7	-8.222	6.522	F3C-CF2-CF2-CF3
33	-21	-18.188	-2.8124	F2HC-CHF2
34	-25	-28.634	3.6339	F2HC-CH3
35	-26.2	-30.827	4.6274	F3C-CH2F
36	-37.5	-32.05	-5.4501	H3C-CH2F
37	-47	-46.807	-0.19339	H3C-CF3
38	22	26.457	-4.4568	FH2C-CH(CH3)2
39	12	8.0471	3.9529	FC(CH3)3
40	21.5	17.311	4.1892	F3C-CH(CH3)-CF3
41	40	38.896	1.1044	F3C-CH(CF3)-CH2F
42	12	14.324	-2.324	F3C-CH(CF3)2
43	-0.3	-2.8219	2.5219	F3C-CF(CF3)2
44	77	72.755	4.2453	FH2C-CH2-CH2-CH2F
45	31	24.239	6.7612	H3C-CF2-CH2-CH3
46	10.45	13.729	-3.2793	FH2C-CH2F
47	25	24.475	0.52489	F2HC-CF2-CH2F
48	40	39.381	0.61881	F3C-CH2-CF2-CH3
49	17	27.505	-10.505	H3C-CF2-CF2-CH3
50	34.3	28.713	5.587	F3C-CHF-CHF-CF3

51	32	28.758	3.2418	FH2C-CHF-CH3
52	66	61.716	4.2836	FH2C-CHF-CH2F
53	18.7	15.386	3.3139	FH2C-CF2-CH3
54	27	35.177	-8.1773	FH2C-CF2-CH2F
55	39	37.214	1.7859	F2HC-CH2-CHF2
56	22	16.857	5.1425	F2HC-CHF-CH3
57	55	47.842	7.1577	F2HC-CHF-CH2F
58	40.5	32.565	7.9347	F2HC-CHF-CHF2
59	-0.8	6.7237	-7.5237	F2HC-CF2-CH3
60	15	17.468	-2.4684	F2HC-CH2-CF3
61	-0.5	-5.8336	5.3336	F3C-CHF-CH3
62	-48.5	-47.81	-0.69021	F3C-CHF2
63	17	22.904	-5.904	F3C-CH2-CH2-CH3
64	0	-1.6153	1.6153	F3C-CF2-CH2F
65	45	38.784	6.216	F2HC-CH2-CH2F
66	14.5	13.168	1.3321	F3C-CF2-CF2-CH3
67	N/A	N/A	N/A	FH2C-CHF-CH2-CH3
68	110	104.63	5.3736	FH2C-CHF-CHF-CH2F
69	78	77.922	0.078154	F2HC-CH(CHF2)-CH2F
70	56.5	59.7	-3.2	F2HC-CH2-CHF-CH3
71	57	57.088	-0.087709	F2HC-CH2-CF2-CH3
72	46.5	48.489	-1.9886	F2HC-CHF-CH2-CH3
73	90	90.153	-0.15337	F2HC-CHF-CHF-CH2F
74	64	74.296	-10.296	F2HC-CF2-CHF-CH2F
75	57.5	58.762	-1.2616	F2HC-CF2-CHF-CHF2
76	53.5	52.786	0.71424	F2HC-CF2-CF2-CH2F
77	59	58.031	0.96883	F2HC-CH(CF3)-CH2F
78	53.5	54.186	-0.68606	F3C-CH2-CH2-CH2F
79	45.5	41.217	4.2827	F3C-CH2-CHF-CH3
80	23	27.411	-4.4108	F3C-CF(CH2F)-CF3
81	42	48.571	-6.5708	F3C-CHF-CF2-CH2F
82	43.5	41.032	2.4682	F3C-CF2-CH2-CH2F
83	41	41.473	-0.47348	H3C-CHF-CHF-CH3

Table 3. Predicted boiling points (by model 2) for 76 fluorobutanes not included in the learning set.

	pred	N_F	structure
1	60.655	2	FH2C-CHF-CH2-CH3
2	57.084	3	FH2C-CH2-CF2-CH3
3	56.527	3	FH2C-CH(CH3)-CHF2
4	74.234	3	FH2C-CHF-CHF-CH3
5	84.396	3	HC(CH2F)3
6	91.5	3	FH2C-CH2-CHF-CH2F
7	72.891	3	FH2C-CH2-CH2-CHF2
8	13.334	4	F3C-CF(CH3)2
9	57.735	4	FH2C-CF(CH3)-CHF2
10	55.436	4	FH2C-CF2-CHF-CH3
11	37.226	4	F2HC-CF2-CH2-CH3
12	27.303	2	F2HC-CH(CH3)2
13	67.773	4	FH2C-CHF-CF2-CH3
14	37.525	4	FH2C-CH(CH3)-CF3
15	26.182	4	F3C-CHF-CH2-CH3
16	54.727	4	F2HC-CH(CH3)-CHF2
17	61.012	4	F2HC-CHF-CHF-CH3
18	84.728	4	FC(CH2F)3
19	76.046	4	FH2C-CF2-CH2-CH2F
20	81.767	4	FH2C-CH(CH2F)-CHF2
21	78.127	4	FH2C-CH2-CHF-CHF2
22	90.422	4	FH2C-CHF-CH2-CHF2
23	57.286	2	H3C-CH(CH2F)2
24	72.219	4	F2HC-CH2-CH2-CHF2
25	12.765	5	H3C-CH2-CF2-CF3
26	36.563	5	FH2C-CF(CH3)-CF3
27	46.804	5	FH2C-CF2-CF2-CH3
28	51.909	5	F2HC-CF(CH3)-CHF2
29	45.708	5	F2HC-CF2-CHF-CH3
30	54.435	5	F2HC-CHF-CF2-CH3
31	35.614	5	F2HC-CH(CH3)-CF3
32	38.588	5	F3C-CHF-CHF-CH3
33	78.235	5	F2HC-CF(CH2F)2
34	59.576	2	FH2C-CH2-CHF-CH3
35	85.333	5	FH2C-CF2-CHF-CH2F
36	66.052	5	F2HC-CF2-CH2-CH2F
37	62.093	5	F3C-CH(CH2F)2
38	55.365	5	F3C-CHF-CH2-CH2F
39	74.77	5	F2HC-CH2-CF2-CH2F
40	71.455	5	F3C-CH2-CHF-CH2F
41	53.27	5	F3C-CH2-CH2-CHF2
42	76.202	5	F2HC-CHF-CH2-CHF2
43	32.82	6	F3C-CHF-CF2-CH3
44	21.114	6	F3C-CF2-CHF-CH3
46	30.614	6	F2HC-CF(CH3)-CF3
47	36.945	6	F2HC-CF2-CF2-CH3
48	56.465	6	F3C-CF(CH2F)2
49	63.937	6	FH2C-CF2-CF2-CH2F
50	56.525	6	F3C-CH2-CF2-CH2F
51	70.555	6	FH2C-CF(CHF2)2

52	63.883	6	F2HC-CF2-CH2-CHF2
53	70.716	6	F2HC-CHF-CF2-CH2F
54	56.962	6	F3C-CH2-CHF-CHF2
55	53.199	6	F3C-CHF-CH2-CHF2
56	33.949	3	F2HC-CF(CH3)2
57	67.226	6	F3C-CHF-CHF-CH2F
58	72.746	6	HC(CHF2)3
59	74.796	6	F2HC-CHF-CHF-CHF2
60	10.139	7	H3C-CF(CF3)2
61	34.269	7	F3C-CHF-CH2-CF3
62	49.162	7	F3C-CF2-CHF-CH2F
63	38.604	7	F3C-CF2-CH2-CHF2
64	45.348	7	F3C-CH2-CF2-CHF2
65	48.594	7	F3C-CF(CH2F)-CHF2
66	61.602	7	FC(CHF2)3
67	35.561	3	H3C-CF2-CHF-CH3
68	52.567	7	F3C-CH(CHF2)2
69	51.596	7	F3C-CHF-CHF-CHF2
70	39.393	8	F3C-CF(CHF2)2
71	33.102	8	F2HC-CH(CF3)2
72	11.224	9	F3C-CF2-CHF-CF3
73	17.926	9	F2HC-CF(CF3)2
74	62.477	3	H3C-CF(CH2F)2
75	45.854	3	FH2C-CF2-CH2-CH3
76	37.003	2	FH2C-CF(CH3)2

Table 4. Experimental (with reliability) and predicted boiling points (extrapolations by model 2), residuals and structures of 18 fluoropentanes.

	bp	rel.	pred	residual	structure
1	45	1	39.974	5.0256	FC(CH ₃) ₂ -CH ₂ -CH ₃
2	47	1	55.931	-8.9313	F ₃ C-CH ₂ -CH ₂ -CH ₂ -CH ₃
3	55	1	60.057	-5.0569	H ₃ C-CHF-CH ₂ -CH ₂ -CH ₃
4	47	1	55.92	-8.9197	H ₃ C-CF ₂ -CF ₂ -CH ₂ -CH ₃
5	75	1	75.356	-0.35596	H ₃ C-CF ₂ -CH ₂ -CF ₂ -CH ₃
6	56	1	57.845	-1.8451	FH ₂ C-CH(CH ₃)-CH ₂ -CH ₃
7	39.5	1	41.699	-2.1994	F ₃ C-CF ₂ -CF ₂ -CH ₂ -CH ₃
8	82	2	79.563	2.437	F ₂ HC-CF ₂ -CF ₂ -CF ₂ -CH ₂ F
9	63	2	73.321	-10.321	FH ₂ C-CH ₂ -CH ₂ -CH ₂ -CH ₃
10	47	1	40.543	6.4565	F ₃ C-CH ₂ -CF(CF ₃) ₂
11	68	1	67.426	0.57367	F ₂ HC-CF ₂ -CF ₂ -CF ₂ -CHF ₂
12	45	2	44.043	0.95673	F ₃ C-CF ₂ -CF ₂ -CF ₂ -CHF ₂
13	54	1	60.32	-6.3204	FH ₂ C-CH ₂ -CH(CH ₃) ₂
14	29	2	20.863	8.1366	F ₃ C-CF ₂ -CF ₂ -CF ₂ -CF ₃
15	30	2	21.463	8.5366	F ₃ C-CF ₂ -CF(CF ₃) ₂
16	41	1	38.573	2.4265	FH ₂ C-C(CH ₃) ₃
17	61	1	53.602	7.3975	H ₃ C-CH ₂ -CF ₂ -CH ₂ -CH ₃
18	60	1	57.546	2.4541	H ₃ C-CF ₂ -CH ₂ -CH ₂ -CH ₃

Figure 1. Plot of calculated (by model 2, white disks, and by leave-one-out crossvalidation, black disks) vs experimental boiling points of 82 fluoromethanes through fluorobutanes.

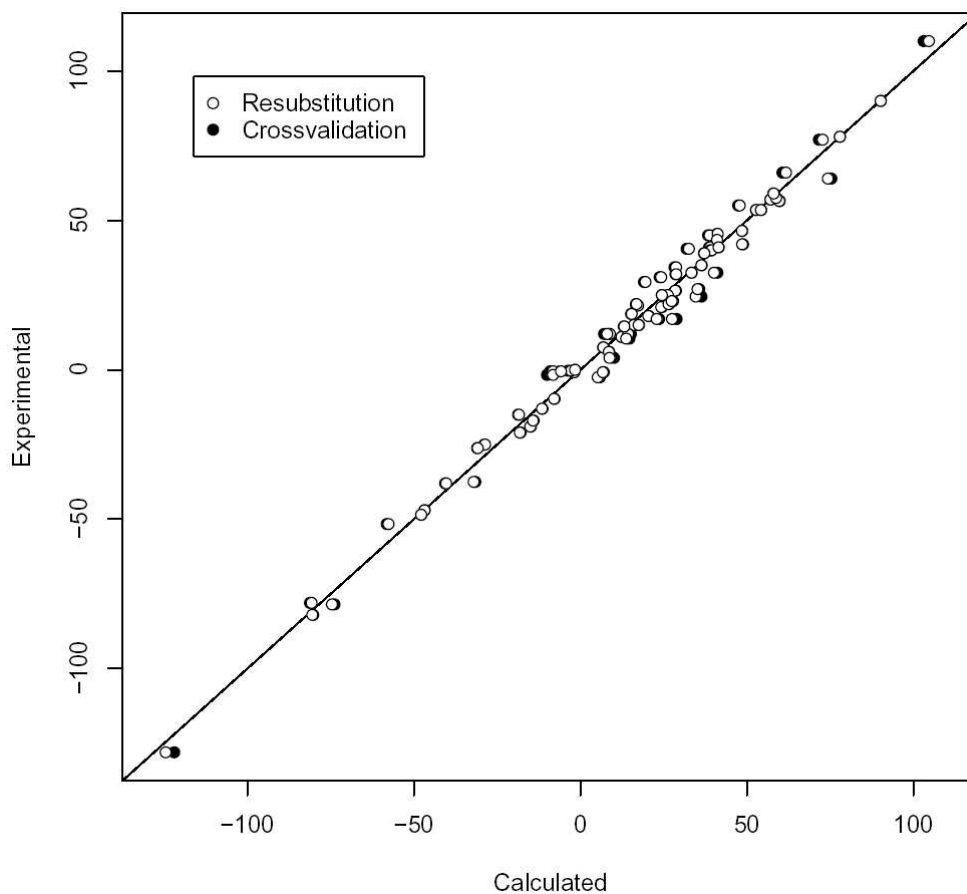


Figure 2. Plot of calculated (by model 2, white disks) and experimental (black disks) boiling points of lower fluoroalkanes vs number of F atoms in the molecule.

