List Operations on Chemical Graphs. 4. Using Edge Models for Prediction of Retention Index Data

Ch. Duvenbeck and P. Zinn*

Lehrstuhl für Analytische Chemie, Ruhr-Universität Bochum, Postfach 102148, D-4630 Bochum 1, Germany

Received July 29, 1992

Previously developed edge models[1] for fitting retention index data are validated by the prediction of this property for three compound classes: monoterpenes, di- and tricyclic methyl esters and alcohols, and monocyclic ketones and alcohols. As a necessary condition, the predicted compounds must have the same bond types as the calibration data set of model development. The resulting prediction errors can be divided into systematic and statistical parts. Systematic errors are based mainly on different experimental conditions between model development and prediction. The statistical part of the prediction error ranges from 15 to 22 retention index units.

INTRODUCTION

In a previous paper,[1] the development of multilinear vertex and edge models for fitting retention index data is discussed. It could be shown that the best fitting results could be obtained using edge models. In order to perform quantitative structure–property relationship (QSPR) studies, many authors describe the fitting of a training data set without validation of the resulting models by prediction of external data that are not involved in model development. In investigating structure retention index relationships, Kaliszan[2] and Oszczapowicz[3] give several examples of experimental and fitted values within sets of training data. A cross-correlation between training data and retention indexes of external data is missing in most cases. By fitting other physical properties, the validation of the developed models is often not to be taken into account also. As examples, Herndon[4] describes the fitting of enthalpic data and Balaban[5] describes the fitting of boiling points.

It is a general experience that fitting physical properties by using multilinear models leads, in the great majority of cases, to excellent results. Only a validation using external data can guarantee whether the found correlation in the training data is a real fact or a chance artifact. Chance correlation results in a dramatic increase of the estimation error predicting properties of external compounds. Wold and Dunn have discussed this problem in detail.[6]

A very informative example for the validation of a multilinear model using external compounds is given by Jurs.[7] The development of models for fitting and prediction of boiling points of furans, tetrahydrofurans, and thiophenes is described. For several compounds the increase of error from fitting to prediction of boiling points is due to differences in experimental conditions in this paper. In other cases, prediction errors may partly be localized in compounds containing structural features that are not incorporated in the compounds of the training set. Therefore, the intention of this paper is to predict the retention indexes in a conservative way only of those external compounds that have the same structural features as the training set compounds. As structural features, the developed edge models use bond types and corresponding topological indexes.[1] The edge models are applied to acyclic and cyclic monoterpenes, some bi- and tricyclic compounds, and monocyclic ketones and alcohols. The bond types of these compounds are all incorporated in the compounds involved in model development.

$$I_k = B_{0k} + \sum_{j=1}^{n} B_{jk}\beta_j^k$$

$\beta_j^1 = M_j$	M-model
$\beta_j^2 = \sum_i ({}^1\xi^l + {}^1\xi^r)_{ij}$	${}^1\xi$-model
$\beta_j^3 = \sum_i (S^l + S^r)_{ij}$	S-model
$\beta_j^4 = \sum_i (R^l + R^r)_{ij}$	R-model

Figure 1. Edge models I_k for the calculation of retention indexes of alkane compounds. B_{jk}'s are the regression coefficients; β_j^k's are the regression parameters based on the atomic level indexes ${}^1\xi$, S, and R. The superscripts l and r mean 'left' and 'right' atoms of a bond.

Comparisons between predicted and corresponding experimental retention indexes will be given. To have a closer look at the resulting errors predicted, data are recalibrated.

PREDICTION OF RETENTION INDEXES WITH EDGE MODELS

The prediction of retention indexes was carried out with the List Operating System described in refs 8 and 9 on an IBM PS2/70 under Golden Common Lisp. The models are shown in Figure 1 and are discussed in more detail in ref 1. They are based on the calculation of atomic level topological indexes that are implicated in the regression parameters β. Beside the numbers of identical bond types M, the atomic level connectivity ${}^1\xi$,[1] the electrotopological state S,[10] and the bonded part R of the electrotopological state[10] are taken into account. The used regression coefficients B are derived from model development with a calibration data set.[1] They are given in Table I, including their standard errors. In order to

0095-2338/93/1633-0220$04.00/0

Table I. Coefficients and Corresponding Standard Errors (SE) of Edge Models Shown in Figure 1 for Prediction of Retention Index Data[a]

	M-model		$^1\xi$-model		S-model		R-model	
coeff type	coeff	SE	coeff	SE	coeff	SE	coeff	SE
ring constant	51.38	3.20	60.59	3.03	54.26	6.18	58.94	6.47
CH_3—CH_2—	*150.00*	*0.00*	*150.00*	*0.00*	45.77	0.50	45.05	0.56
CH_3—CH<	120.88	0.87	55.27	0.40	41.58	0.38	40.05	0.43
CH_3—C<	102.24	0.98	42.13	0.40	39.42	0.37	38.11	0.45
—CH_2—CH_2—	*100.00*	*0.00*	*100.00*	*0.00*	33.27	0.30	33.15	0.31
—CH_2CH<	73.40	1.00	33.04	0.42	25.21	0.75	24.89	0.72
—CH_2—C<	67.74	1.59	28.73	0.58	24.90	1.20	22.89	1.17
>CH—CH<	68.11	2.71	33.62	0.96	27.95	1.96	23.47	1.83
>CH—C<	76.17	3.73	36.09	1.17	39.76	3.20	26.93	2.84
OH—CH_2—	303.55	3.75	159.12	1.91	47.93	0.41	45.17	0.58
OH—CH<	225.56	3.46	107.87	1.68	45.00	0.36	39.62	0.53
OH—C<	190.76	4.13	80.62	1.76	45.88	0.42	41.93	0.65
O=C<	169.42	4.60	149.64	19.46	36.58	0.95	23.19	4.99
—O-C<	23.91	4.67	−14.53[b]	8.27	2.49[b]	1.70	9.39[b]	5.68
—O—CH_3	174.15	3.87	93.33	2.07	25.56	0.69	29.38	0.86
—O—CH_2—	95.72	2.60	46.20	1.22	20.93	0.44	20.43	0.57
—O—CH<	36.27	6.23	18.11	2.59	15.95	1.19	14.05	1.43
>C—CH_3	152.39	1.51	30.12	10.06	39.74	3.27	56.95	9.52
>C—CH_2—	94.99	2.38	13.38[b]	8.05	24.64	4.44	42.38	11.95
>C—CH<	67.35	3.71	10.54[b]	6.73	17.67	5.94	34.11	14.52
>C—C<	57.09	5.98	10.21[b]	6.05	135.10	44.70	25.61[b]	20.36
=CH—CH_3	102.13	8.84	1.27[b]	4.55	24.22	2.00	23.76	2.50
=CH—CH_2—	36.43	8.31	−14.21	3.17	6.47	2.51	6.30	2.98
=CH—CH<	1.20[b]	9.85	−19.39	3.32	−5.44[b]	3.45	−5.28[b]	3.99
=CH—C<	1.26[b]	10.47	−11.75	3.06	−6.16[b]	4.14	−6.89[b]	4.82
CH_2=CH—	199.53	9.21	138.34	4.11	42.88	1.53	42.46	2.08
—CH=CH—	212.91	18.01	171.72	7.49	55.40	4.08	56.92	5.06
>C=CH_2	86.49	4.67	108.75	19.46	30.77	4.37	6.00[b]	14.19
>C=CH—	104.20	9.59	141.12	17.42	39.16	6.90	7.27[b]	19.09
>C=C<			120.24	28.63	15.69[b]	16.19	−62.33[b]	44.64

[a] The last coefficient belonging to the *M*-model could not be calculated because of a linear combination in the training data set.[1] [b] The indicated coefficients are of low significance.

Table II. Monoterpenes Used for Prediction of Retention Indexes *I*[est.] Using *M*-Model[a]

substance	T/lit.	I[lit.]	I[est.]	I[corr.]	substance	T/lit.	I[lit.]	I[est.]	I[corr.]
2,6-dimethyloctane	100 (14)	938	932.8	957.9	α-terpineol	130 (11)	1185	1157.6	1157.5
3,7-dimethyl-1,6-octadiene	90 (13)	946	940.4	964.7	tetrahydrogeraniol	130 (11)	1185	1186.4	1183.0
2,6-dimethyl-2-octene	80 (13)	966	963.1	984.8	β-citronellol	130 (11)	1215	1216.7	1209.9
trans-p-menthane	100 (14)	981	975.7	996.0	β-nerol	130 (11)	1218	1237.0	1227.9
cis-p-menthane	100 (14)	995	975.7	996.0	geraniol	130 (11)	1243	1237.0	1227.9
limonene	130 (11)	1030	992.4	1010.8	isopulegyl acetate	130 (11)	1258	1322.2	1303.6
γ-terpinene	130 (11)	1057	984.1	1003.4	lavandulol	130 (11)	1274	1339.9	1319.3
6,10-dihydromyrcenol	130 (11)	1063	1058.0	1069.0	isomenthyl acetate	130 (12)	1283	1325.8	1306.8
terpinolene	100 (12)	1081	1034.2	1047.9	neoisomenthyl acetate	130 (12)	1297	1325.8	1306.8
tetrahydrolinalool	130 (11)	1087	1093.6	1100.7	citronellyl acetate	130 (11)	1335	1354.5	1332.3
tetrahydromyrcenol	130 (11)	1090	1080.7	1089.1	β-neryl acetate	130 (11)	1345	1374.9	1350.3
linalool	130 (11)	1092	1106.9	1112.5	citronellyl propionate	130 (11)	1427	1447.1	1414.5
1,2-dihydrolinalool	130 (11)	1122	1123.9	1127.5	β-neryl propionate	130 (11)	1436	1467.5	1432.5
β-terpineol	130 (11)	1137	1132.9	1135.5	citronellyl isobutyrate	130 (11)	1469	1511.2	1471.4
menthone	130 (11)	1143	1134.1	1136.6	β-neryl isobutyrate	130 (11)	1474	1531.6	1489.4
isopulegol	130 (11)	1145	1165.8	1164.7	geranyl isobutyrate	130 (11)	1493	1531.6	1489.4
lavandulol	130 (11)	1154	1202.0	1196.9	citronellyl butyrate	130 (11)	1511	1547.1	1503.3
isomenthone	130 (12)	1156	1134.1	1136.6	*cis*-nerolidol	130 (11)	1524	1595.0	1545.7
neomenthol	120 (13)	1159	1169.4	1167.9	geranyl butyrate	130 (11)	1532	1567.5	1521.3
δ-terpineol	130 (11)	1160	1146.1	1147.2	*trans*-nerolidole	130 (11)	1553	1595.0	1545.7
menthol	130 (11)	1171	1169.4	1167.9	β-neryl isopentanoate	130 (11)	1574	1632.6	1579.1
terpinen-4-ol	130 (11)	1175	1172.0	1170.3	geranyl isopentanoate	130 (11)	1593	1632.6	1579.1
neoisomenthol	130 (12)	1180	1169.4	1167.9	citronellyl pentanoate	130 (11)	1608	1647.1	1592.0
carvomenthone	110 (13)	1181	1134.1	1136.6	geranyl pentanoate	130 (11)	1632	1667.5	1610.1
isomenthol	130 (12)	1182	1169.4	1167.9					

[a] *I*[lit.] = retention indexes from literature; *I*[corr.] = retention indexes corrected by recalibration procedure; *T* = temperature (°C); lit. = reference no.

substitute the regression coefficients for the CH_3–CH_2- and -CH_2–CH_2- bonds by their theoretical values corresponding to Kovats' retention index definition, the *M*- and $^1\xi$-models are extended by the parameter $\Delta = 150M_{CH_3\text{-}CH_2\text{-}} + 100M_{\text{-}CH_2\text{-}CH_2\text{-}}$ as shown in Figures 4, 7, and 9.

A further purpose of this paper is to compare the four different edge models. It will be investigated if more sophisticated models that are based on the atomic level topological indexes are advantageous with respect to the incremental *M*-model counting only identical bond types.

MONOTERPENES

The edge models were used to predict the retention index data of acyclic and cyclic monoterpenes shown in Figures 2

Table III. Comparison of Mean Absolute Errors (E_{MAE}) Corresponding to Prediction of Retention Index Data with Recalibration[a]

class	errors and coeffs	*M*-model	$^1\xi$-model	*S*-model	*R*-model
cyclic and acyclic monoterpenes	E_{MAE} (prediction)	26.3	26.0	27.7	30.5
	E_{MAE} (recalibration)	15.3	15.3	22.4	18.1
	k_0	129.8	128.9	128.7	155.7
	k_1	0.8877	0.8891	0.9013	0.8688
di- and tricyclic methyl esters and alcohols[b]	E_{MAE} (prediction)	60.8	63.4	52.7	50.0
	E_{MAE} (recalibration)	19.4	20.0	21.8	18.5
	k_0	449.6	483.8	334.3	416.2
	k_1	0.6039	0.5786	0.7445	0.6441
monocyclic ketones and alcohols	E_{MAE} (prediction)	31.7	30.8	31.9	30.9
	E_{MAE} (recalibration)	16.7	14.8	18.6	18.1
	k_0	78.9	70.6	38.6	82.9
	k_1	0.9406	0.9517	0.9845	0.9314

[a] k_0 and k_1 are intercepts and slopes of the recalibration plots. The four edge models and the three prediction classes are considered. [b] Compounds **2** and **7** of Figure 6 are not considered.

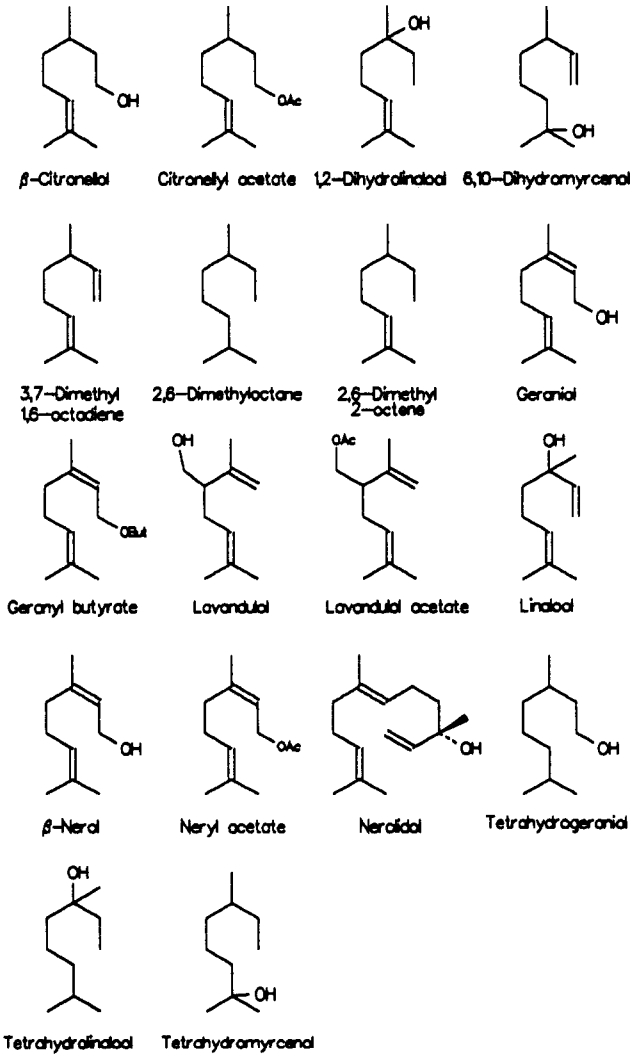

Figure 2. Skeletons of acyclic monoterpenes as external compounds for the prediction of retention indexes. Further acyclic monoterpenes are derivatives of the illustrated compounds.

and 3. Most experimental values are referring to Jennings.[11] The rest of the data is taken from different references given in Table II. It should be mentioned that the stationary phases for the gas chromatographic measurements were of nonpolar character. Typical phases for these measurements were SE30, SF96, OV1, OV101, BPI, CPSIL5CB, SP2100, DB5, and HPI while only Squalane and OV1 were involved for the development of the models.[1] The temperatures affecting model development have been in a range of 50–60 °C while the terpene retention indexes from the literature were measured between 80 and 130 °C.

The comparison of the predicted retention indexes of the monoterpenes with data from the literature is shown in Figure 4. The mean absolute errors for the prediction of the 49 compounds lie in a range of 26–30 retention index units corresponding to the used model. Fitting the calibration data set for model development, these errors vary between 7 and 9 units.[1] A recalibration by simple linear regression between predicted and literature values decreases the statistical error down to 15 units for the *M*- and $^1\xi$-models. Figure 5 shows the recalibration plots. The differences to the pure prediction errors without recalibration and the deviations of the regression slopes from '1' as shown in Table III indicate systematical errors. The already mentioned different experimental conditions between data for model development and terpene retention indexes from literature may be reasons for systematic errors. It is also important that the prediction of the terpene indexes has a character of an extrapolation. The retention indexes for model development range between 400 and 1000 and those for prediction between 900 and 1600. Overall, the resulting predicted retention indexes of monoterpenes seem to be acceptable. As a necessary condition for these predictions, all bond types of the monoterpenes are involved in the model development.

DI- AND TRICYCLIC COMPOUNDS

Another group of compounds including the same bond types as used in model development are the di- and tricyclic methyl esters and alcohols of Figure 6. The retention indexes of these compounds were measured by Heintz et al.[15] on SE-30 as stationary phase at 150 °C. Most of these compounds have norbornane or related skeletons. The comparison of predicted retention indexes with the data from ref 15 is shown in Figure 7. The two outliers are corresponding to **2** and **7** from Figure 6. **2** is already determined as outlying by Kaliszan,[2] while **7** has an adamantane skeleton so that a misprediction becomes plausible. All other 29 compounds are predicted with a mean absolute error in a range of 50–63 retention index units. Stereoisomerism is not considered as documented in data pairs with identical predicted values. The recalibration of these 29 compounds reduces the mean absolute errors to a range of 18–22 units. As documented in Table III, the slope of the prediction plot is extremely different from '1'. This indicates again systematic errors corresponding to different experimental conditions affecting model development and prediction. The systematic errors may also be based on the structural

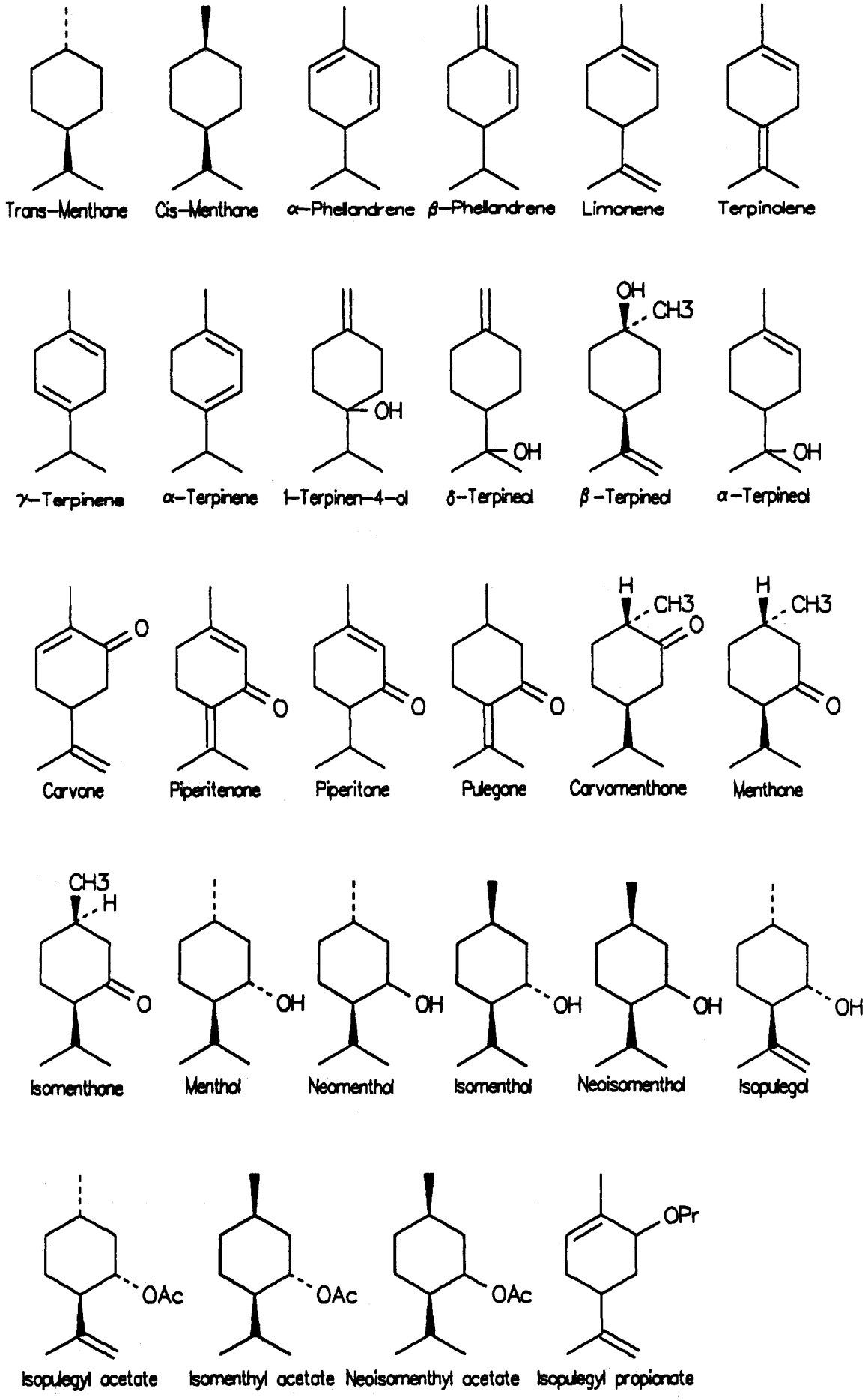

Figure 3. Skeletons of the cyclic monoterpenes as external compounds for the prediction of retention indexes. The stereoisomerism of several compounds is not considered for the prediction.

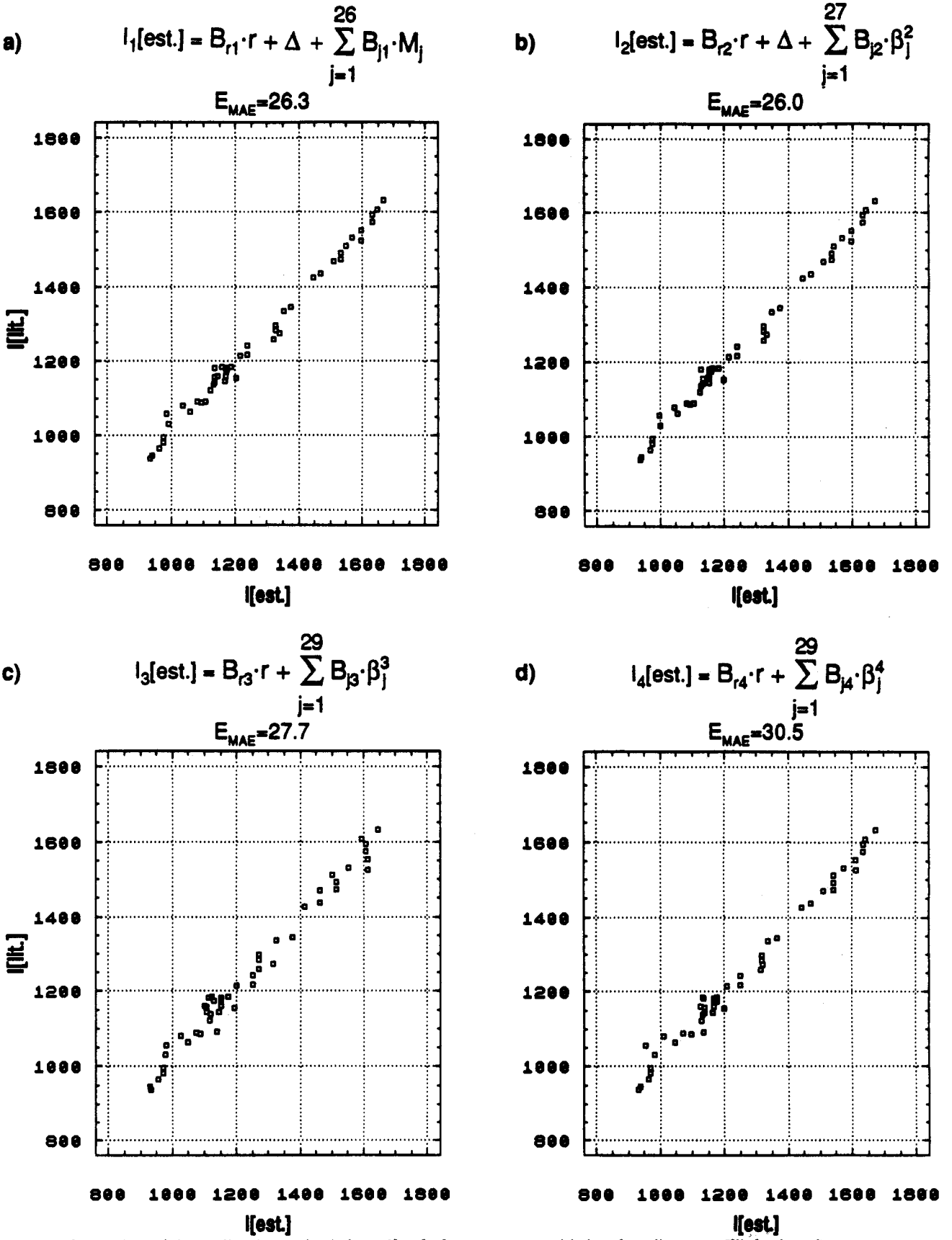

Figure 4. Comparison of the predicted retention indexes I[est.] of monoterpenes with data from literature I[lit.] using edge models of Figure 1. E_{MAE} = mean absolute error.

complexity of the prediction molecules. As parameters describing the molecular shape, only the ring constants B_r of the models[1] are taken into account. In the case of di- and tricyclic molecules, a simple ring constant is obviously not suited to approximate the more complex molecular shapes. The predicted, recalibrated, and experimental retention indexes of this compound class are shown in Table IV.

With respect to the importance of model validation by prediction of external compounds, the result of the pure data fitting is shown in Figure 8. The fitting model considers six

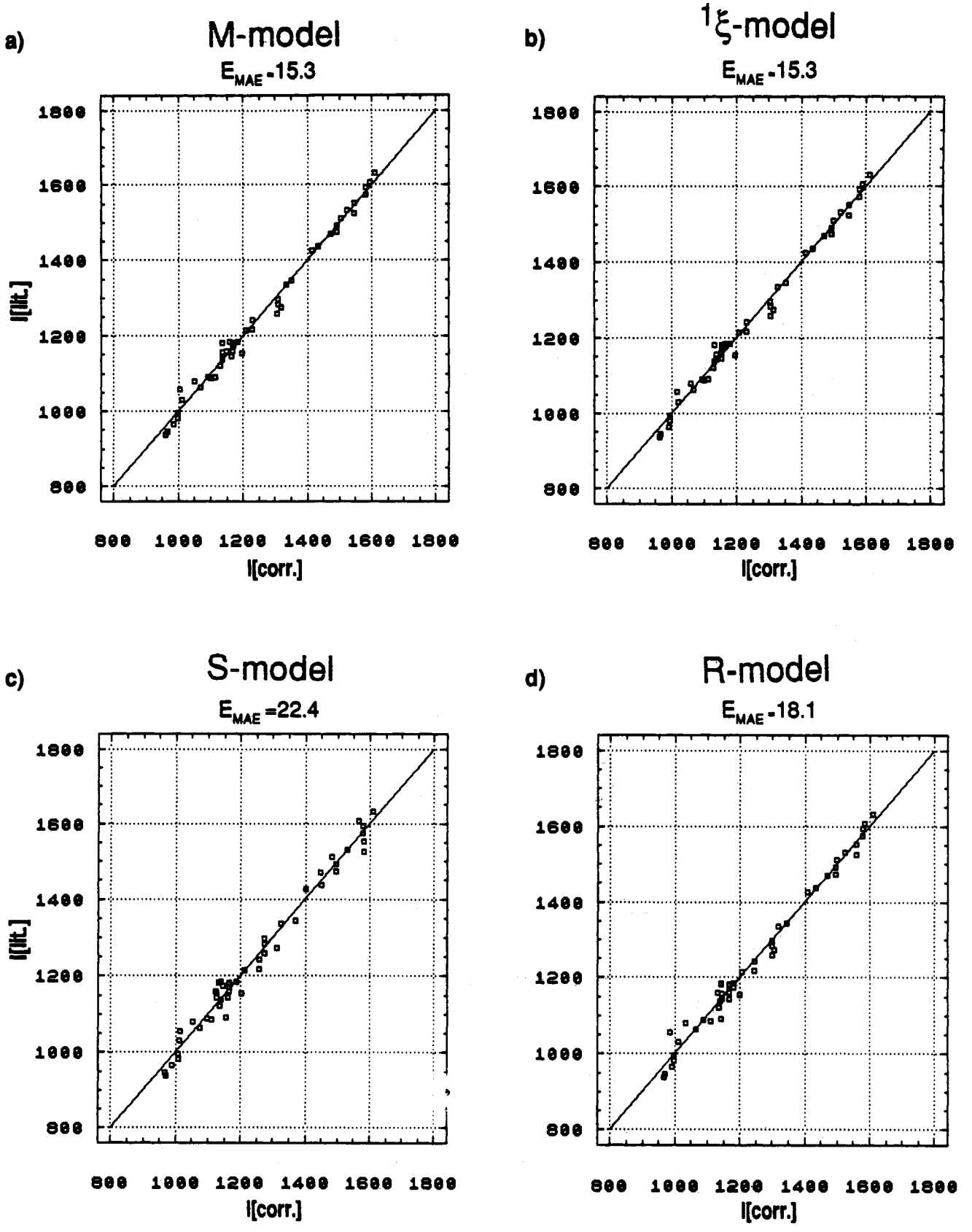

Figure 5. Plot of recalibrated retention indexes I[corr.] of monoterpenes versus data from literature I[lit.] corresponding to the underlying edge models of Figure 1. E_{MAE} = mean absolute error.

alkane, four oxygen, and three double bond containing bond types. This model leads to a mean absolute error of 5.3 retention index units. The comparison of the small fitting with the large prediction errors apparently favors the fitting procedure. On the other hand, the fitting describes only this small limited data set and has no robust coefficients to predict external compounds. This is a typical example of a chance correlation as mentioned in the Introduction.

MONOCYCLIC KETONES AND ALCOHOLS

The retention indexes of 22 monocyclic ketones and alcohols from Sadtler Catalogue[16] were compared with their predicted

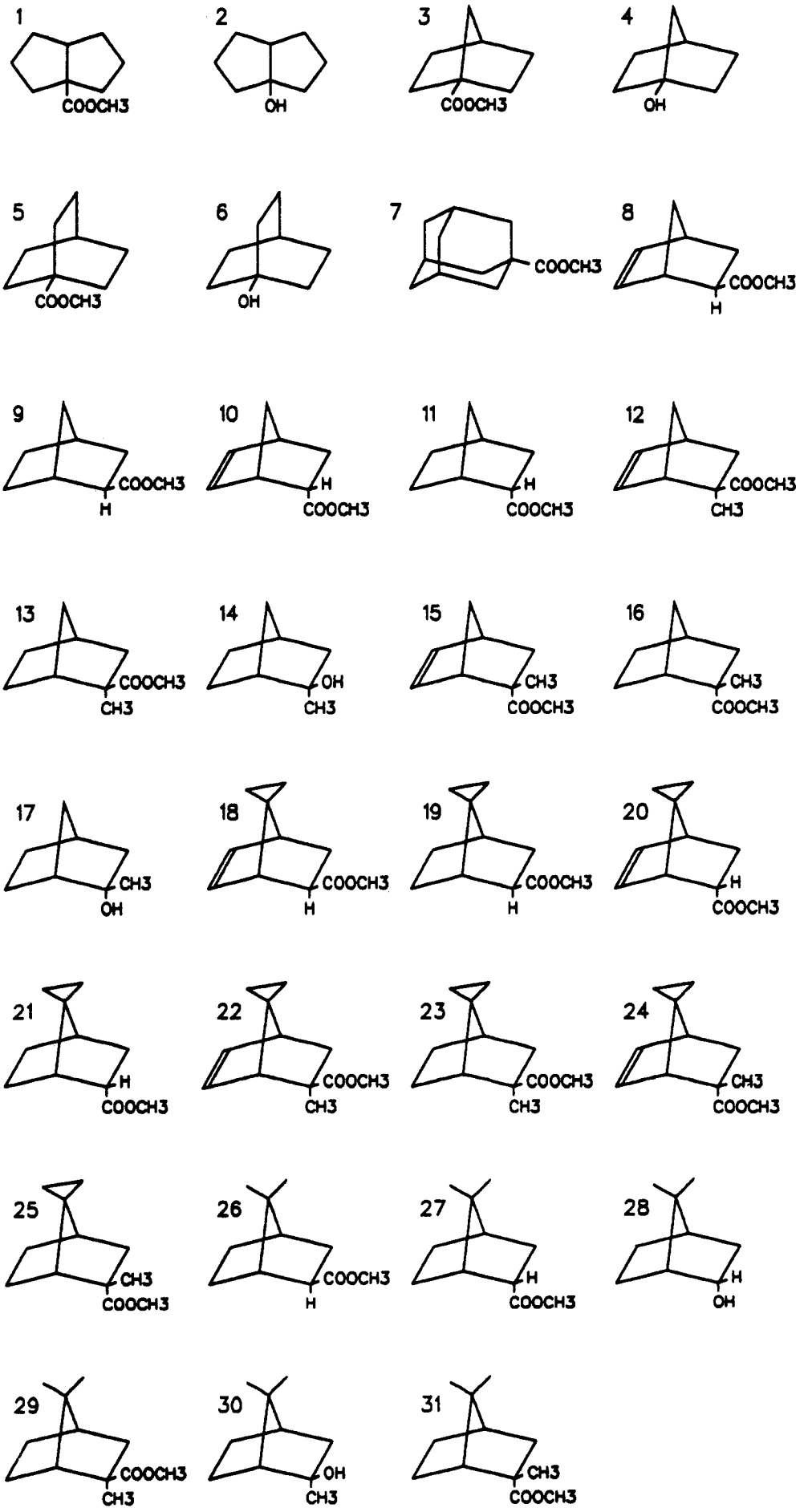

Figure 6. Skeletons of the bi- and tricyclic compounds from ref 15 as external compounds for the prediction of retention indexes.

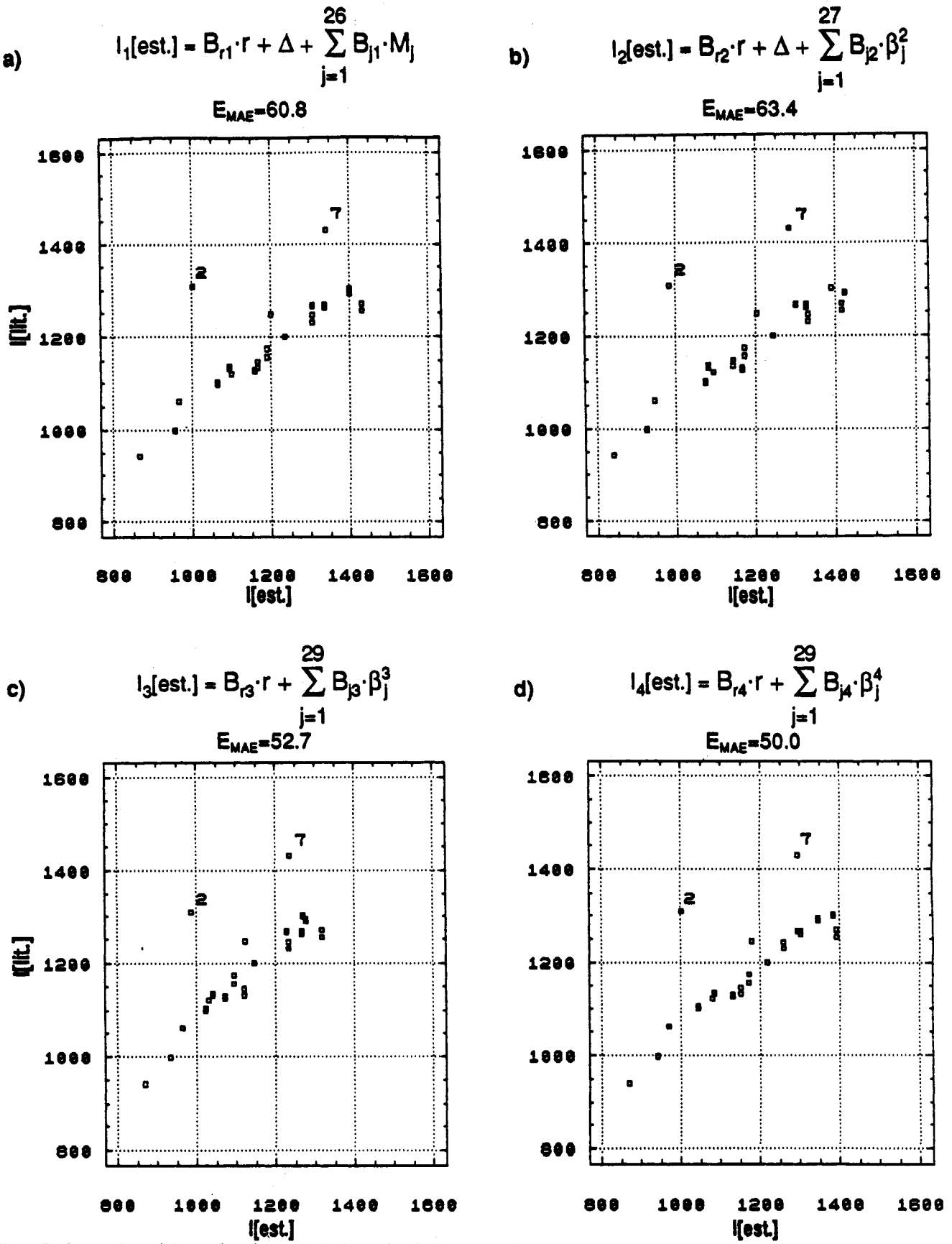

Figure 7. Comparison of the *predicted* retention indexes I[est.] of the bi- and tricyclic compounds shown in Figure 6 with data from literature I[lit.]. The outlying compounds 2 and 7 are not taken into account. E_{MAE} = mean absolute error.

values. The compounds and retention indexes are summarized in Table V. Figure 9 shows the corresponding plots. In this case, the experimental conditions involved in model development and prediction are nearly the same. The stationary phase was OV1, and the temperature was 60 °C. The mean absolute errors of the models vary in a range of 27–33 retention index units. The recalibration reduces these errors to values between 15 and 19. The coefficients of recalibration regression

Table IV. Experimental *I*[lit.], Predicted *I*[est.] (*M*-Model), and Recalibrated *I*[corr.] Retention Indexes of Bi- and Tricyclic Compounds Corresponding to Figure 6

substance	*I*[lit.]	*I*[est.]	*I*[corr.]
4	942	865.6	972.3
14	998	955.3	1026.5
17	999	955.3	1026.5
6	1061	965.6	1032.7
8	1098	1063.2	1091.7
10	1103	1063.2	1091.7
3	1120	1099.4	1113.5
15	1125	1157.6	1148.7
12	1130	1157.6	1148.7
11	1131	1094.7	1110.7
29	1133	1165.3	1153.3
9	1135	1094.7	1110.7
31	1146	1165.3	1153.3
13	1156	1189.1	1167.7
16	1174	1189.1	1167.7
1	1200	1234.4	1195.1
18	1230	1304.3	1237.3
20	1245	1304.3	1237.3
5	1247	1199.4	1173.9
22	1254	1430.1	1313.3
19	1262	1335.7	1256.3
27	1265	1304.7	1237.5
26	1268	1304.7	1237.5
21	1268	1335.7	1256.3
24	1270	1430.1	1313.3
23	1291	1398.6	1294.2
25	1295	1398.6	1294.2
28	1302	1399.1	1294.5
30	1303	1399.1	1294.5
2	1310	1000.6	
7	1431	1339.8	

are summarized in Table III. The slopes of these monocyclic compounds k_1 are lying near the expected values of '1'. This indicates similar experimental conditions between model development and prediction, as already mentioned. The systematical error is mainly affected by the model constants B_r. B_r was derived from a data set including only cyclic alkanes as cyclic compounds, whereas the prediction data set is composed of cyclic ketones and alcohols. The molecular shape of these compounds differ from simple alkanes. The introduction of molecular shape indexes as described by Kier and Hall[17] should improve the predictive power of the edge models in this case.

CONCLUSIONS

The prediction of retention indexes makes it obvious that the estimation errors are consisting of two parts: systematic and statistical errors. Typical systematic errors are those resulting from different experimental conditions between model development and prediction or incomplete models not considering the molecular shape. The recalibration of the predicted retention indexes by simple linear regression approximates the pure statistical error. As descriptors of the statistical error, the mean absolute errors of the recalibrations are ranging from 15 to 22 retention index units for all three predicted compound classes. This statistical error may incorporate further systematic parts of error as unconsidered stereoisomerism.

Comparing the resulting errors of the prediction with those of the fitted calibration data set used in model development that vary between 7 and 9[1] edge models seems to be a way toward the prediction of retention index data on the basis of molecular structure.

The comparison of the prediction errors of the four different edge models (*M*-, $^1\xi$-, *S*-, and *R*-model) gives no significant preference to one of them. This is in accordance with the previously discussed[1] fitting of retention index data also using these four edge models. As structural features for the

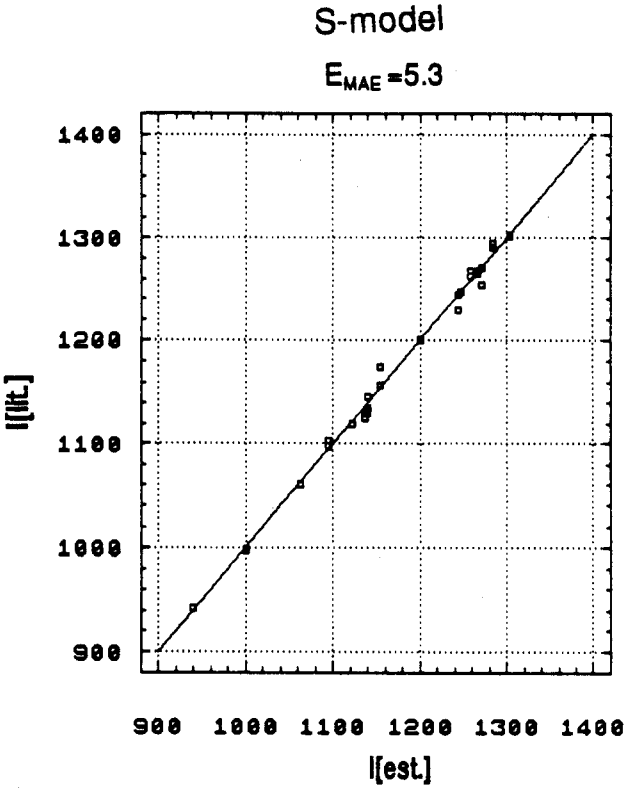

Figure 8. Comparison of the *fitted* retention indexes *I*[est.] of the bi- and tricyclic compounds shown in Figure 6 with data from the literature *I*[lit.] as an example for a chance correlation. The derived model coefficients are of no predictive value! E_{MAE} = mean absolute error.

Table V. Experimental *I*[lit.], Predicted *I*[pred.] (*M*-Model), and Recalibrated *I*[corr.] Retention Indexes of Cyclic Ketones and Alcohols as External Compounds

substance	*I*[lit.]	*I*[est.]	*I*[corr.]
cyclopentanone	748.8	710.8	747.4
3-methylcyclopentanone	810.9	778.5	811.1
cyclohexanone	850.8	810.8	841.5
2-methylcyclohexanone	911.3	877.4	904.2
3-methylcyclohexanone	911.9	878.5	905.2
4-methylcyclohexanone	917.2	878.5	905.2
cycloheptanone	968.7	910.8	935.6
4-ethylcyclohexanone	1028.6	981.0	1001.6
2-propylcyclohexanone	1098.1	1079.9	1094.7
2-butylcyclohexanone	1200.0	1179.9	1188.7
cyclopentanol	758.1	723.7	759.6
cyclohexanol	857.0	823.7	853.7
1-methylcyclohexanol	868.2	879.9	906.5
1-ethylcyclopentanol	873.5	895.4	921.1
cis-4-methylcyclohexanol	922.0	891.4	917.4
trans-4-methylcyclohexanol	922.8	891.4	917.4
2,5-dimethylcyclohexanol	965.2	980.4	1001.1
1-propylcyclopentanol	968.0	995.4	1015.1
3,5-dimethylcyclohexanol	972.0	959.1	981.0
cycloheptanol	989.0	923.7	947.8
1-methylcycloheptanol	996.2	979.9	1005.5
4-ethylcyclohexanol	1031.6	994.0	1013.8

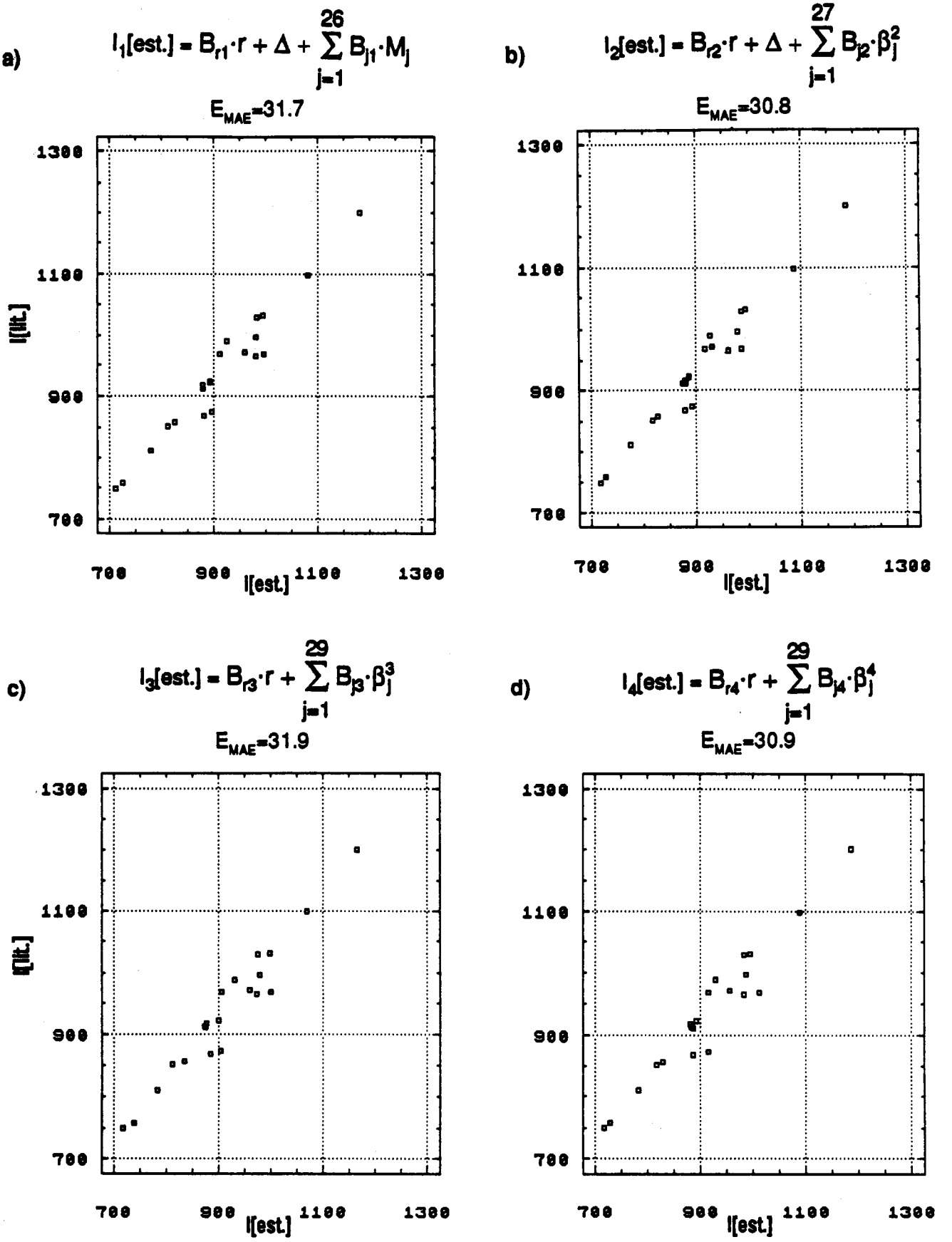

Figure 9. Comparison of the predicted retention indexes I[est.] of the cyclic alcohols and ketones summarized in Table V with data from literature I[lit.]. E_{MAE} = mean absolute error.

description of molecular retention indexes, it is sufficient to consider the incorporated bond types. Using atomic or bond level topological indexes to refine the models does not decrease the estimation errors. Because of its smallest computation time, the pure incremental M-model will be the favorite for future work.

REFERENCES AND NOTES

(1) Duvenbeck, C.; Zinn, P. List Operations on Chemical Graphs. 3. Development of Vertex and Edge Models for Fitting Retention Index Data. *J. Chem. Inf. Comput. Sci.*, preceding paper in this issue.

(2) Kaliszan, R. *Quantitative Structure–Chromatographic Retention Relationships*; John Wiley & Sons: New York, 1987.

(3) Oszczapowicz, J. Additivity Rules and Correlation Methods in Gas Chromatography. In *Similarity Models in Organic Chemistry, Biochemistry and Related Fields*; Zalewski, R. I., Krygowski, T. M., Shorter, J., Eds.; Elsevier: Amsterdam, 1991.

(4) Garbalena, M.; Herndon, W. C. Optimum Graph-Theoretical Models for Enthalpic Properties of Alkanes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 27–42.

(5) Balaban, A. T.; Kier, L. B.; Joshi, N. Correlations between Chemical Structure and Boiling Points of Acyclic Ethers, Peroxides, Acetals, and Their Sulfur Analogues. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 237–244.

(6) Wold, S.; Dunn, W. J. Multivariate quantitative structure–activity relationships (QSAR): conditions for their applicability. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 6–13.

(7) Stanton, D. T.; Jurs, P. C.; Hicks, M. G. Computer-Assisted Prediction of Normal Boiling Points of Furans, Tetrahydrofurans, and Thiophenes. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 301–310.

(8) Gautzsch, R.; Zinn, P. List Operations on Chemical Graphs. 1. Basic List Structures and Operations. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 541–550.

(9) Gautzsch, R.; Zinn, P. List Operations on Chemical Graphs. 2. Combining Basic List Operations. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 551–555.

(10) Hall, L. H.; Mohney, B.; Kier, L. B. The Electrotopological State: Structure Information at the Atomic Level for Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 76–82.

(11) Jennings, W.; Shibamoto, T. *Qualitative Analysis of Flavour and Fragrance Volatiles by Glass Capillary Column Gas Chromatography*; Academic Press: New York, 1980.

(12) Malingre, T. M.; Maarse, H. Composition of the Essential Oil of Mentha Aquatica. *Phytochemistry* **1974**, *13*, 1531–1535.

(13) Swigar, A. A.; Silverstein, R. M. *Monoterpenes: Infrared, Mass, Proton, and C-13 NMR-Spectra, and Kovats' Indexes*; Aldrich Chemical Co.: Milwaukee, WI, 1981.

(14) Kepner, R. E.; Maarse, H. Hydrogenolysis of Terpenes in the Injection Port of a Gas Chromatograph. I. Monoterpenes. *J. Chromatogr.* **1972**, *66*, 229–237.

(15) Heintz, M.; Gruselle, M.; Druilhe, A.; Lefort, D. *Chromatographia* **1976**, *9*, 367.

(16) *The Sadtler Standard Gas Chromatography Retention Index Library*; Sadtler Research Laboratories: Philadelphia, PA, 1985.

(17) Hall, C. H.; Kier, L. B. The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure–Property Modeling. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1991; Vol. 2.