



Canadian Journal of  
**Chemistry**

Français | [Return to table of contents](#) | [Journal Home](#)

**Related links**

[Full-text HTML](#)

[Full-text PDF](#)

[Supplementary data](#)

[Reprints and Permissions](#)

**Canadian access to full text made available through the  
Depository Services Program**

Can. J. Chem. **87**(11): 1593–1604 (2009) | doi:10.1139/V09-109 | Published  
by NRC Research Press / Publié par les Presses scientifiques du CNRC

**QSPR studies on normal boiling points and molar  
refractivities of organic compounds by correlation-  
ranking-based PCR and PC–ANN analyses of new  
topological indices**

**Raouf Ghavami, Amir Najafi, and Bahram Hemmateenejad**

**Abstract:** The new topological indices (Sh indices) based on the distance sum and connectivity of a molecular graph, previously developed by our team, were extended to predict the two physicochemical properties, including normal boiling point (NBP) and molar refractivity (MR), of a large set of organic compounds consisting of alkanes, alkenes, ethers, amines, alcohols, alkylbenzenes, and alkylhalides. The sets of molecular descriptors were derived directly from the two-dimensional molecular structure of the compounds based on graph theory. Both linear and nonlinear modelings were implemented by using principal component regression (PCR) and principal component – artificial neural network (PC–ANN) with back-propagation learning algorithm, respectively. Eigenvalue and correlation-ranking procedures were used to rank the principal components and entered them into the models. Principal component analysis of Sh data matrix showed that the respective six and seven PCs could explain 97.49% and 99.22% of variances in the Sh indices. PCR analysis of the NBP and MR data demonstrated that the proposed Sh indices could explain about 97.52% and 99.52% of variations, while the variations explained by the PC–ANN modeling were more than 99.00% and 99.82%, respectively. The predictive ability of the models were evaluated using an external test set for NBP and MR of the molecules with the respective root-mean-square errors lower than 9.69 K and 0.660 cm<sup>3</sup> mol<sup>-1</sup> for the linear model and 6.17 K and 0.416 cm<sup>3</sup> mol<sup>-1</sup> for the nonlinear model.

*Key words:* topological Sh indices, eigenvalue ranking, correlation ranking, normal boiling point, molar refractivity, principal component analysis, principal component regression (PCR), principal component – artificial neural network (PC–ANN).

**Résumé :** Les nouveaux indices topologiques (indices Sh) basés sur la somme de la distance et la connectivité d'un graphe moléculaire et développés antérieurement par notre équipe ont été étendus afin de pouvoir prédire deux propriétés physico-chimiques, dont le point d'ébullition normal (PEN) et la réfractivité molaire (RM) d'un large ensemble de composés organiques formés d'alcanes, d'alcènes, d'amines, d'alkylbenzènes et d'halogénures d'alkyle. Les ensembles de

descripteurs moléculaires ont été dérivés directement de la structure moléculaire bidimensionnelle des composés basée sur la théorie des graphes. On a développé des modèles linéaires et non linéaires en utilisant l'algorithme d'apprentissage en rétropropagation respectivement avec une régression du composant principal (RCP) et avec le réseau neural artificiel du composant principal (RNA-CP). On a fait appel aux méthodes de la valeur de eigen et du classement des corrélations pour classer les composants principaux qui ont ensuite été introduits dans les modèles. L'analyse du composant principal d'une matrice de données Sh a permis de montrer que respectivement six ou sept composants principaux (CP) peuvent expliquer 97,49 % et 99,22 % des variances dans les indices Sh. L'analyse de la RCP des données relatives aux PEN et aux RM a permis de démontrer que les indices Sh peuvent expliquer environ 97,52 % et 99,52 % des variations alors que les variations expliquées par me modèle RNA-CP expliquent respectivement 99,00 % et 99,82 % des variations. On a évalué l'habilité des modèles à faire des prédictions relatives aux PEN et aux RM des molécules en utilisant un ensemble externe et les erreurs quadratiques moyennes respectives étaient inférieures réseau neural artificiel du composant principal (RNA-CP) à 9,69 K et 0,660 cm<sup>3</sup> m<sup>-1</sup> pour le modèle linéaire et de 6,17 K et 0,416 cm<sup>3</sup> m<sup>-1</sup> pour le modèle non linéaire.

*Mots-clés* : indices SH topologiques, classement de valeurs de eigen, classement de corrélation, point d'ébullition normal (PEN), réfractivité moléculaire (RM), analyse du composant principal, régression du composant principal (RCP), réseau neural artificiel du composant principal (RNA-CP).

[Traduit par la Rédaction]

## References

(1) Herndon, W. C.; Nowak, P. C.; Connor, D. A.; Lin, P. J. Am. Chem. Soc. 1992, **114** (1), 41–47.



(2) Hall, L. H.; Kier, L. B. J. Chem. Inf. Comput. Sci. 1995, **35**, 1039.

(3) Cash, G. G. J. Chem. Inf. Comput. Sci. 1995, **35**, 815.

(4) Estrada, E. J. Chem. Inf. Comput. Sci. 1995, **35**, 31.

(5) Yuan, H.; Cao, C. J. Chem. Inf. Comput. Sci. 2003, **43** (2), 501–512.



(6) Ren, B. J. Mol. Struct. THEOCHEM 2002, **586** (1-3), 137–148.



(7) Lorentz, H. A. Wied. Ann. Phys. 1880, **9**, 641.

(8) Lorentz, L. V. Wied. Ann. Phys. 1880, **11**, 70.

(9) Gladstone, J. H.; Dale, T. P. Philos. Trans. R. Soc. Lond. A 1858, **148** (-1), 887–894.



(10) Vogel, A. J. Chem. Soc. 1948, 1833.



(11) Norrington, F. E.; Hyde, R. M.; Williams, S. G.; Wooten, R. J. *Med. Chem.* 1975, **18** (6), 604–607.

[CrossRef](#)

(12) Ravi, M.; Hopfinger, A. J.; Hormann, R. E.; Dinan, L. J. *Chem. Inf. Comput. Sci.* 2001, **41** (6), 1587–1604.

[MEDLINE](#)

(13) Luke, B. T. J. *Mol. Struct. THEOCHEM* 1999, **468** (1-2), 13–20.

[CrossRef](#)

(14) Bruneau, P. J. *Chem. Inf. Comput. Sci.* 2001, **41** (6), 1605–1616.

[MEDLINE](#)

(15) Katritzky, A. R.; Petrukhin, R.; Tatham, D.; Basak, S.; Benfenati, E.; Karelson, M.; Maran, U. J. *Chem. Inf. Comput. Sci.* 2001, **41** (3), 679–685.

[MEDLINE](#)

(16) Katritzky, A. R.; Gordeeva, E. V. J. *Chem. Inf. Comput. Sci.* 1993, **33** (6), 835–857.

[MEDLINE](#)

(17) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. *Chem. Rev.* 1996, **96** (3), 1027–1044.

[CrossRef](#) [MEDLINE](#)

(18) Wiener, H. J. *Am. Chem. Soc.* 1947, **69** (1), 17–20.

[CrossRef](#)

(19) Hosoya, H. *Bull. Chem. Soc. Jpn.* 1971, **44** (9), 2332–2339.

[CrossRef](#)

(20) Balaban, A. T. *Chem. Phys. Lett.* 1982, **89** (5), 399–404.

[CrossRef](#)

(21) Balaban, A. T. *Pure Appl. Chem.* 1983, **55** (2), 199–206.

[CrossRef](#)

(22) Randić, M. J. *Math. Chem.* 1991, **7** (1), 155–168.

[CrossRef](#)

(23) Nelson, S. D.; Seybold, P. G. J. *Mol. Graph. Model.* 2001, **20** (1), 36–53.

[CrossRef](#)

(24) Liu, S.; Liu, H.; Xia, Z.; Cao, C.; Li, Z. J. *Chem. Inf. Comput. Sci.* 1999, **39**, 951.

(25) Liu, S.; Cai, S.; Cao, C.; Li, Z. J. *Chem. Inf. Comput. Sci.* 2001, **40**, 1337.

(26) Haaland, D. In *Computer-Enhanced Analytical Spectroscopy*; Jurs, P. C., Ed.; Plenum Press: New York, 1992; Vol. 3.

(27) Wold, S.; Sjostrom, M.; Eriksson, L. *Chemom. Intell. Lab. Syst.* 2001, **58** (2), 109–130.

[CrossRef](#)

(28) Tang, T.; Li, T. *Chemom. Intell. Lab. Syst.* 2002, **64** (1), 55–64.

[CrossRef](#)

(29) Liu, S.; Zhang, R.; Liu, M.; Hu, Z. *J. Chem. Inf. Comput. Sci.* 1997, **37**, 1146.

(30) Walczak, B. *Anal. Chim. Acta* 1996, **322** (1-2), 21–29.

[CrossRef](#)

(31) Shamsipur, M.; Hemmateenejad, B.; Akhond, M. *J. AOAC Int.* 2002, **85** (3), 555–562.

[MEDLINE](#)

(32) Kim, B.; Park, S. *Chemom. Intell. Lab. Syst.* 2002, **56** (1), 39–50.

[CrossRef](#)

(33) Tomás-Vert, F.; Perez-Gimenez, F.; Salvador, M. T.; Garcia-March, F. J.; Jaen-Oltra, J. *J. Mol. Struct. THEOCHEM* 2000, **504** (1-3), 249–259.

[CrossRef](#)

(34) Peterson, K. L. *J. Chem. Inf. Comput. Sci.* 1995, **35** (5), 896–904.

[MEDLINE](#)

(35) Shamsipur, M.; Hemmateenejad, B.; Akhond, M. *Anal. Chim. Acta* 2002, **461** (1), 147–153.

[CrossRef](#)

(36) Agatonovic-Kustrin, S.; Zecevic, M.; Zivanovic, L.; Tucker, I. G. *J. Pharm. Biomed. Anal.* 1998, **17** (1), 69–76.

[CrossRef](#) [MEDLINE](#)

(37) Gemperline, P. J.; Long, J. R.; Gregoriou, V. G. *Anal. Chem.* 1991, **63** (20), 2313–2323.

[CrossRef](#)

(38) González, M. P.; Toropov, A. A.; Duchowicz, P. R.; Castro, E. A. *Molecules* 2004, **9** (12), 1019–1033.

[CrossRef](#) [MEDLINE](#)

(39) Jäntschi, L.; Bolboacă, S.-D. *Int. J. Mol. Sci.* 2007, **8** (3), 189–203.

[CrossRef](#)

(40) Jäntschi, L.; Bolboacă, S.-D. *Leonardo Electronic Journal of Practices and Technologies* 2005, **7**, 55.

(41) Qian, C.; Hongyi, Z. *Chem. J. Internet* 2006, **8** (9), 58.

(42) Balaban, A. T.; Mills, D.; Basak, S. C. *J. Chem. Inf. Comput. Sci.* 1999, **39**, 758.

(43) Toropov, A. A.; Toropova, A. P. *J. Mol. Struct. THEOCHEM* 2002, **581** (1-3), 11–15.

[CrossRef](#)

(44) Mukhamedzhanova, D. V.; Gutman, I. *Indian J. Chem. A* 2005, **44**, 1545.

(45) Shamsipur, M.; Hemmateenejad, B.; Akhond, M. *Bull. Korean Chem. Soc.* 2004, **25**, 253.

(46) Shamsipur, M.; Ghavami, R.; Hemmateenejad, B.; Sharghi, H. *QSAR Comb. Sci.* 2004, **23** (9), 734–753.

[CrossRef](#)

- (47) Shamsipur, M.; Ghavami, R.; Hemmateenejad, B.; Sharghi, H. *Internet Electron. J. Mol. Des.* 2005, **4**, 882.
- (48) Shamsipur, M.; Hemmateenejad, B.; Ghavami, R.; Sharghi, H. *Pol. J. Chem.* 2007, **81**, 269.
- (49) Kier, L. B.; Hall, L. H. In *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
- (50) David, R. L., Ed. *CRC Handbook of Chemistry and Physics*, 81st ed.; CRC Press, 2000–2001.
- (51) Malinowski, E. R. In *Factor Analysis in Chemistry*; Wiley: New York, 2002; pp 224–225.
- (52) Sutter, J. M.; Kalivas, J. H.; Lang, P. M. *J. Chemometr.* 1992, **6** (4), 217–225.  
[CrossRef](#)
- (53) Hemmateenejad, B. *J. Chemometr.* 2004, **18** (11), 475–485.  
[CrossRef](#)
- (54) Hemmateenejad, B. *Chemom. Intell. Lab. Syst.* 2005, **75** (2), 231–245.  
[CrossRef](#)
- (55) Hawkins, D. M.; Basak, S. C.; Mills, D. J. *Chem. Inf. Comput. Sci.* 2003, **43** (2), 579–586.  
[MEDLINE](#)
- (56) Hawkins, D. M. *J. Chem. Inf. Comput. Sci.* 2004, **44** (1), 1–12.  
[MEDLINE](#)
- (57) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. *Nature* 1986, **323** (6088), 533–536.  
[CrossRef](#)
- (58) Svozil, D.; Kvasnicka, V.; Pospichal, J. *Chemom. Intell. Lab. Syst.* 1997, **39** (1), 43–62.  
[CrossRef](#)
- (59) Devillers, J.; Domine, D.; Guillon, C. *Eur. J. Med. Chem.* 1998, **33** (7-8), 659–664.  
[CrossRef](#)
- (60) Taskinen, J.; Yliruusi, J. *Adv. Drug Deliv. Rev.* 2003, **55** (9), 1163–1183.  
[CrossRef](#) [MEDLINE](#)
- (61) Depczynski, U.; Frost, V. J.; Molt, K. *Anal. Chim. Acta* 2000, **420** (2), 217–227.  
[CrossRef](#)
- (62) Shamsipur, M.; Ghavami, R.; Sharghi, H.; Hemmateenejad, B. *J. Mol. Graph. Model.* 2008, **27** (4), 506–511.  
[CrossRef](#)