

## Statistical properties of spectra of chloronaphthalenes

Dorota Bielińska-Wąż · Piotr Wąż ·  
Timothy Clark · Tomasz Puzyn ·  
Łukasz Peplowski · Wiesław Nowak

Received: 15 September 2012 / Accepted: 15 November 2012 / Published online: 7 December 2012  
© The Author(s) 2012. This article is published with open access at Springerlink.com

**Abstract** We introduce spectral density distribution moments as molecular descriptors. We demonstrate that these descriptors correctly represent the molecular structure. To prove the correctness of the new descriptors, we consider the IR spectra of 76 chloronaphthalenes. We show that the consecutive substitutions of the hydrogen atoms by the chlorine atoms are reflected by a regular behavior of the descriptors and, consequently, that the descriptors may be used to construct classification diagrams. The diagrams allow for a clear distinction between the compounds containing different numbers of the substituents. In the diagrams, symbols corresponding to the isomers with a given number of chlorine atoms nearly overlap and form clusters. The clusters corresponding to different numbers of chlorine atoms are well separated and distributed in a regular way. The same kind of analysis we also use as a tool for checking the correctness of the calculated spectra.

---

D. Bielińska-Wąż (✉)  
Instytut Fizyki, Uniwersytet Mikołaja Kopernika, Grudziądzka 5, 87-100 Toruń, Poland  
e-mail: dsnake@fizyka.umk.pl

P. Wąż  
Centrum Astronomii, Uniwersytet Mikołaja Kopernika, Gagarina 11, 87-100 Toruń, Poland

T. Clark  
Computer-Chemie-Centrum and Interdisciplinary Center for Molecular Materials,  
Friedrich-Alexander-Universität Erlangen-Nürnberg,  
Nägelsbachstrasse 25, 91052 Erlangen, Germany

T. Puzyn  
Pracownia Chemometrii Środowiska, Wydział Chemii, Uniwersytet Gdański, Sobieskiego 18/19,  
80-952 Gdańsk, Poland

Ł. Peplowski · W. Nowak  
Instytut Fizyki, Uniwersytet Mikołaja Kopernika, Grudziądzka 5, 87-100 Toruń, Poland

**Keywords** Statistical spectroscopy · Descriptors · Molecular similarity · DFT methods

## 1 Introduction

Similarity of molecular structures is highly relevant to various realms of science such as toxicology, ecotoxicology, and pharmacology. The basic paradigm of quantitative structure-property/activity relationships (QSPR/QSAR) is that compounds with similar structure have similar properties. This implies a smooth transient behavior in the relation between structure and property/activity, i.e., for any small change in the structure, the magnitude of a physico-chemical property or biological activity changes smoothly rather than in an abrupt, in an all-or-none type, way. Recently, much attention has been paid to so-called “activity cliffs” [1], which imply that this smooth behavior may not always be appropriate. However, we simply point out here that exactly such discontinuous behavior would be expected to result from descriptors that do not encode the relevant molecular properties adequately. This behavior is, for instance, well known when trying to define geometrical reaction coordinates for complex reactions. The concept of molecular similarity is always complicated by ambiguities in the representation of molecular structure and in the definition of similarity. In the computation of molecular similarity, a large number of mathematical functions can be used to derive measures of similarity for a pair of molecules starting from the same set of structural descriptors [2, 3]. Quantitative molecular similarity analysis (QMSA) uses global molecular descriptors, such as 2D-fingerprints [4] or topological indices [5], to identify molecular similarity for property prediction and for environmental risk assessment. However, similarity measures that concentrate on the relevant regions of molecules, such as established 3D QSAR methods like comparative molecular field analysis (CoMFA) [6], are often more appropriate. These methods rely on spatially resolved, rather than global molecular similarity. They are potentially extremely powerful, but suffer from the disadvantage that they depend on the molecular conformation used, so that extensive conformational searches are necessary. This problem has been treated using conformationally weighted descriptors [7], which implies a dynamic description of flexible molecules. Methods are currently being developed [8] to use 3D information efficiently in drug design. A promising class of spatially resolved methods in computational pharmacology is derived from quantum-mechanical descriptions of molecules and of their potential interactions with the environment [9–12].

Recently, we have proposed intensity distribution moments as molecular descriptors [13–15]. They are related to the shapes of molecular spectra. We have shown that using new descriptors one can distinguish nitriles from amides [16].

In the present work, we consider another kind of distribution moments—spectral density distribution moments. We demonstrate the usefulness of these descriptors using an example of the infrared spectra of 76 chloronaphthalenes.

Chloronaphthalenes have been used in industry for many years. They are toxic and of considerable interest in environmental studies [17]. Naphthalene IR spectrum, as well as spectra of other polyaromatic hydrocarbons, are also of great interest in astrophysics [18, 19].

We show that the new descriptors correctly represent the molecular structure—they clearly identify the number of chlorine atoms in the compounds. Moreover, we use statistical analysis as a tool for checking the correctness of the calculated spectra. We have used similar methodologies in astrophysics [20,21]. We have shown that using methods of statistical spectroscopy it is possible to check the correctness of the existing classification of the stellar spectra. This methodology seems to be universal, and can be applied to many kinds of problems involving different classification schemes, as for example distinguishing between chaotic and periodic motions [22] or the classification of stars. We have also used this methodology to study the similarity of DNA sequences [23].

## 2 Theory

Let us consider a discrete frequency spectrum  $\nu_1, \nu_2, \dots, \nu_D$ . A common approach in statistical spectroscopy is to describe this spectrum by the following distribution:

$$\rho(\nu) = \frac{1}{D} \sum_{i=1}^D \delta(\nu - \nu_i), \quad (1)$$

where  $\delta(\nu - \nu_i)$  is the Delta Dirac function. The function  $\rho(\nu)$  represents the density of the frequencies. This function is called the spectral density distribution and is normalized

$$\int_{-\infty}^{\infty} \rho(\nu) d\nu = 1. \quad (2)$$

The aim of this work is to introduce spectral density distribution moments  $M_{\rho,q}, M'_{\rho,q}, M''_{\rho,q}$  as molecular descriptors. The  $q$ -th moment of  $\rho(\nu)$  is defined as

$$M_{\rho,q} = \int_{-\infty}^{\infty} \nu^q \rho(\nu) d\nu. \quad (3)$$

Using Eq. 1, the  $q$ -th spectral density distribution moment reads

$$M_{\rho,q} = \frac{1}{D} \sum_{i=1}^D \nu_i^q. \quad (4)$$

The corresponding  $q$ -th spectral density scaled moments are

$$M'_{\rho,q} = \frac{1}{D} \sum_{i=1}^D (\nu_i - M_{\rho,1})^q, \quad (5)$$

$$M''_{\rho,q} = \frac{1}{D} \sum_{i=1}^D \left[ \frac{(v_i - M_{\rho,1})}{\sqrt{M_{\rho,2} - (M_{\rho,1})^2}} \right]^q \quad (6)$$

In the present work, we construct spectral density distributions  $\rho(\nu)$  from the frequencies  $\nu_i$  of the infrared (IR) spectra. The moments of these distributions supply information about the locations of  $\nu_i$  but not about the intensities. They describe various properties of the distribution. In particular  $M_{\rho,1}$  is the mean frequency,  $M'_{\rho,2}$  describes the width,  $M''_{\rho,3}$ —the asymmetry, and  $M''_{\rho,4}$ —the excess of this distribution. Higher-order moments do not have direct geometrical equivalents.

We show that spectral density distribution moments provide information also about the molecular structure. For this purpose we construct the classification diagrams (see the subsequent section).

### 3 Results and discussion

We perform the calculations for 76 compounds: chloronaphthalenes containing from zero through eight chlorine atoms. They are listed in Table 1, where  $r = 0, 1, \dots, 75$  are the labels of the compounds. Several examples of the compounds are shown in Fig. 1.

We study spectral density distributions of the frequencies of the IR spectra of the chloronaphthalenes.

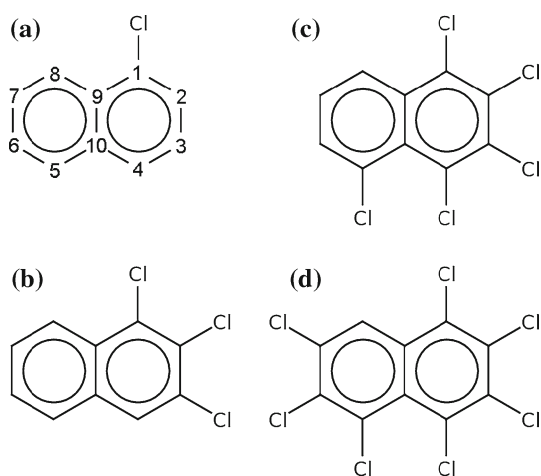
The vibrational spectra we obtained from systematic high level DFT (Density Functional Theory) calculations. A hybrid B3LYP functional and 6-311++G\*\* basis were

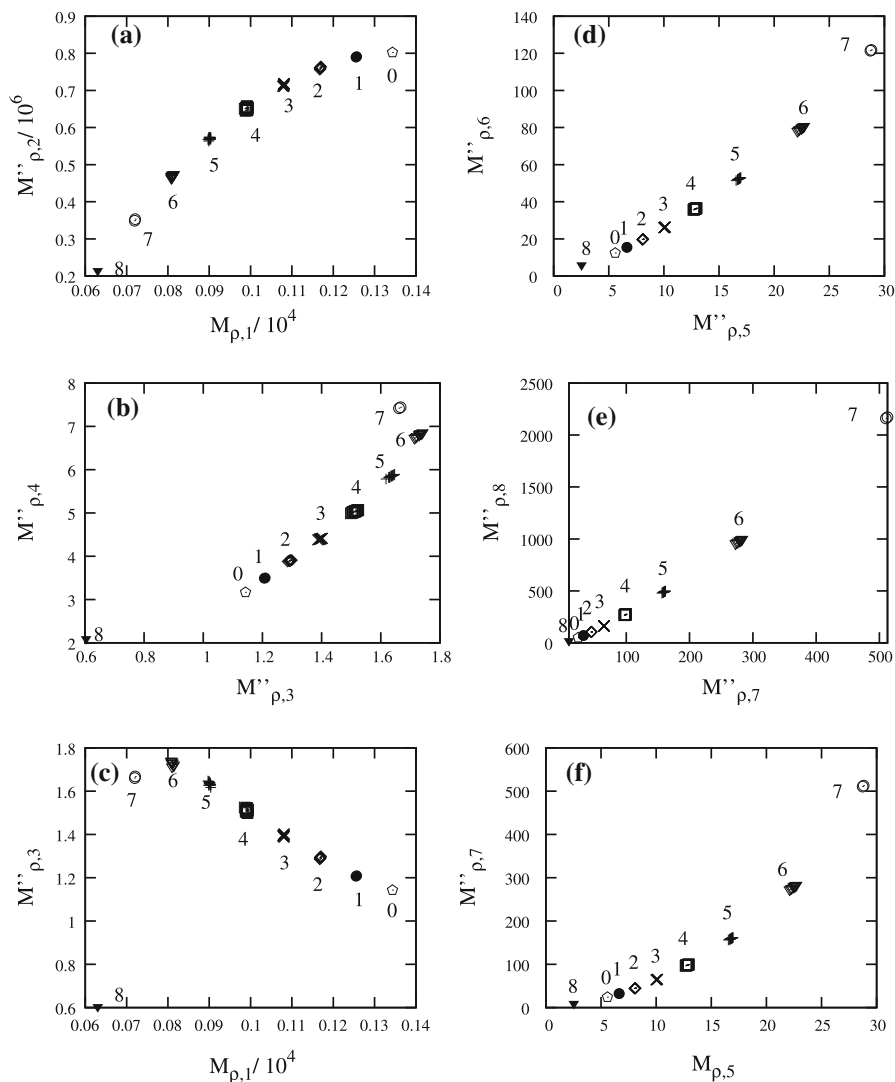
**Table 1** Compounds

r		r	
0	Naphthalene	38	1,2,5,8-tetraCN
1	1-CN	39	1,2,6,7-tetraCN
2	2-CN	40	1,2,6,8-tetraCN
3	1,2-diCN	41	1,2,7,8-tetraCN
4	1,3-diCN	42	1,3,5,7-tetraCN
5	1,4-diCN	43	1,3,5,8-tetraCN
6	1,5-diCN	44	1,3,6,7-tetraCN
7	1,6-diCN	45	1,3,6,8-tetraCN
8	1,7-diCN	46	1,4,5,8-tetraCN
9	1,8-diCN	47	1,4,6,7-tetraCN
10	2,3-diCN	48	2,3,6,7-tetraCN
11	2,6-diCN	49	1,2,3,4,5-pentaCN
12	2,7-diCN	50	1,2,3,4,6-pentaCN
13	1,2,3-triCN	51	1,2,3,5,6-pentaCN
14	1,2,4-triCN	52	1,2,3,5,7-pentaCN
15	1,2,5-triCN	53	1,2,3,5,8-pentaCN

**Table 1** continued

r		r	
16	1,2,6-triCN	54	1,2,3,6,7-pentaCN
17	1,2,7-triCN	55	1,2,3,6,8-pentaCN
18	1,2,8-triCN	56	1,2,3,7,8-pentaCN
19	1,3,5-triCN	57	1,2,4,5,6-pentaCN
20	1,3,6-triCN	58	1,2,4,5,7-pentaCN
21	1,3,7-triCN	59	1,2,4,5,8-pentaCN
22	1,3,8-triCN	60	1,2,4,6,7-pentaCN
23	1,4,5-triCN	61	1,2,4,6,8-pentaCN
24	1,4,6-triCN	62	1,2,4,7,8-pentaCN
25	1,6,7-triCN	63	1,2,3,4,5,6-hexaCN
26	2,3,6-triCN	64	1,2,3,4,5,7-hexaCN
27	1,2,3,4-tetraCN	65	1,2,3,4,5,8-hexaCN
28	1,2,3,5-tetraCN	66	1,2,3,4,6,7-hexaCN
29	1,2,3,6-tetraCN	67	1,2,3,5,6,7-hexaCN
30	1,2,3,7-tetraCN	68	1,2,3,5,6,8-hexaCN
31	1,2,3,8-tetraCN	69	1,2,3,5,7,8-hexaCN
32	1,2,4,5-tetraCN	70	1,2,3,6,7,8-hexaCN
33	1,2,4,6-tetraCN	71	1,2,4,5,6,8-hexaCN
34	1,2,4,7-tetraCN	72	1,2,4,5,7,8-hexaCN
35	1,2,4,8-tetraCN	73	1,2,3,4,5,6,7-heptaCN
36	1,2,5,6-tetraCN	74	1,2,3,4,5,6,8-heptaCN
37	1,2,5,7-tetraCN	75	OctaCN

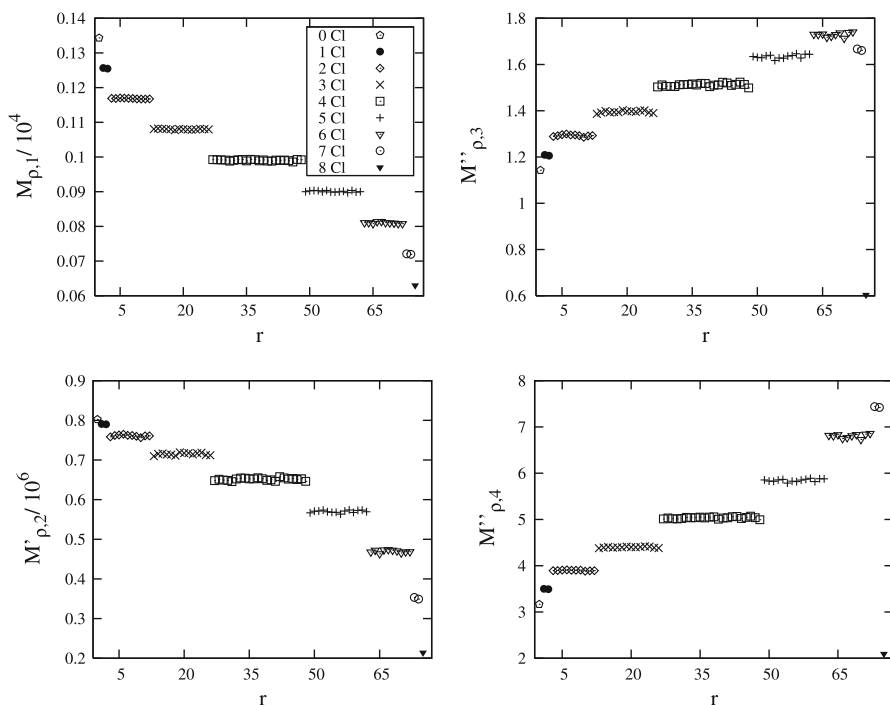
**Fig. 1** Compounds: **a**  $r = 1$ ;  
**b**  $r = 13$ ; **c**  $r = 49$ ; **d**  $r = 73$ 



**Fig. 2** Classification diagrams for 76 compounds based on spectral density distribution moments

used as implemented in the Gaussian 03 code [24]. A careful geometry optimization was performed prior to calculation of vibrational spectra for each compound. In some cases an initial guess of planar chloronaphthalene geometry led to transition states or saddle points, so in order to avoid these artefacts a lower, non-planar initial guess geometries were used. Thus, no calculated minima had imaginary frequencies.

Figure 2 shows classification diagrams based on the descriptors defined in Eqs. 4, 5, 6. The descriptors representing compounds with the same number of chlorine atoms are denoted by the same symbols in the plots. The number of chlorine atoms are written next to the symbols representing the corresponding descriptors. For



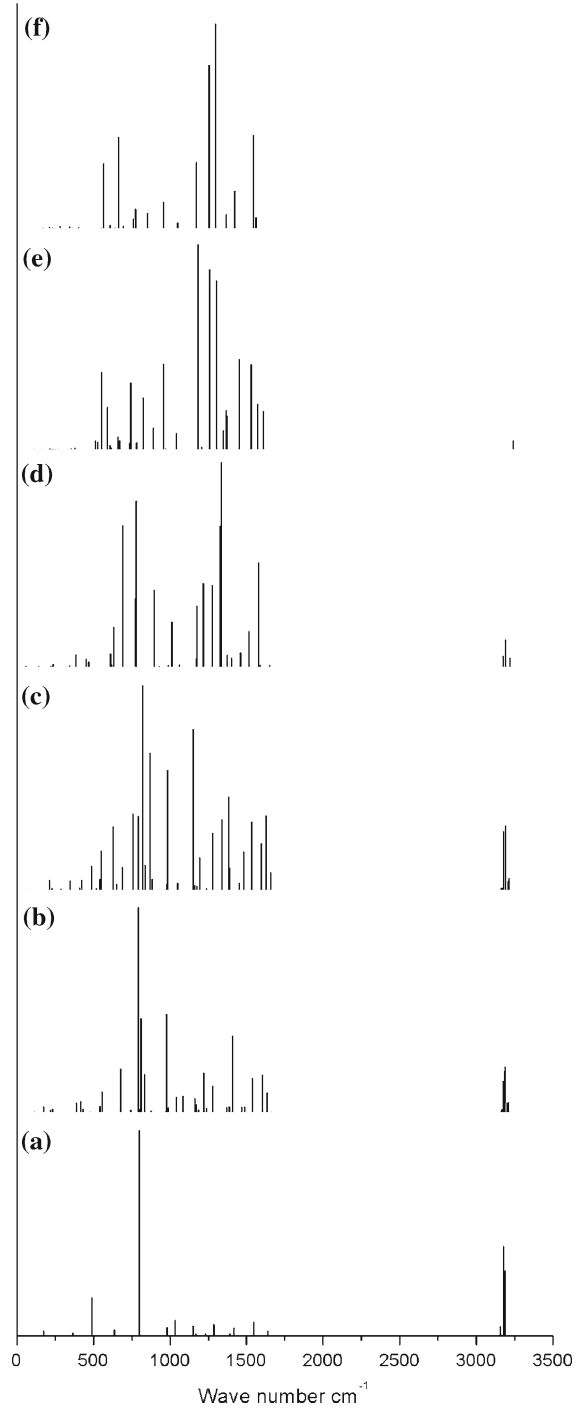
**Fig. 3** Spectral density distribution moments for 76 compounds

example, the descriptors representing the 22 compounds with four chlorine atoms ( $r = 27, 28, \dots, 48$ ) are denoted by squares. All the squares in the figures are clustered—they are nearly in one point of the diagram. We observe that all the descriptors representing compounds with the same number of chlorine atoms are located in the same parts of the classification diagrams. This kind of behavior is observed in all the cases:  $M_{\rho,1} - M'_{\rho,2}$  (Fig. 2a),  $M''_{\rho,3} - M''_{\rho,4}$  (Fig. 2b),  $M_{\rho,1} - M''_{\rho,3}$  (Fig. 2c),  $M''_{\rho,5} - M''_{\rho,6}$  (Fig. 2d),  $M''_{\rho,7} - M''_{\rho,8}$  (Fig. 2e),  $M''_{\rho,5} - M''_{\rho,7}$  (Fig. 2f).

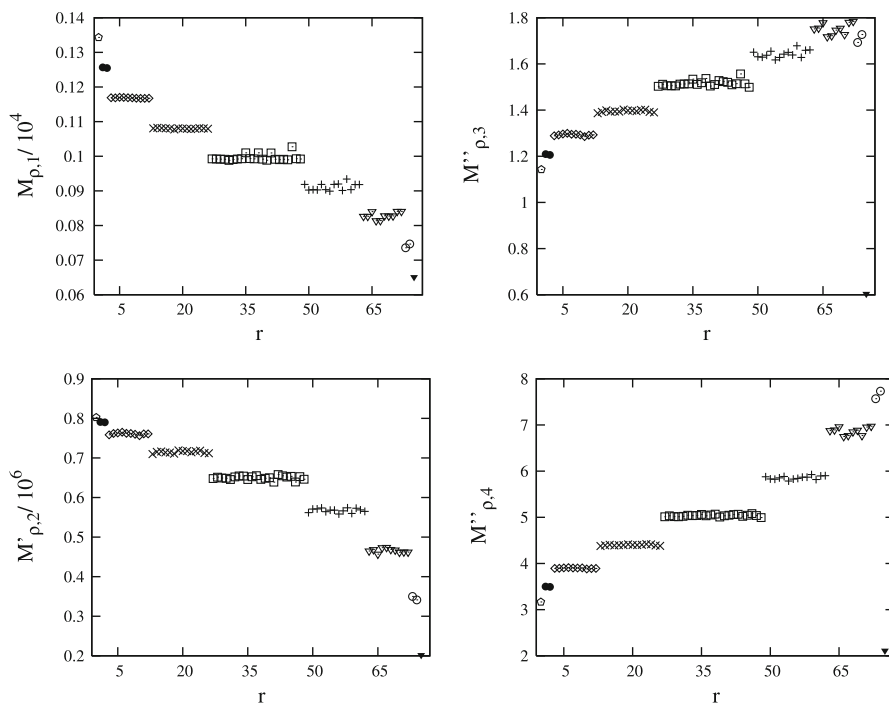
Figure 3 shows the first four moments of the spectral density distributions for all 76 compounds. In this figure, one can recognize particular descriptors corresponding to particular compounds numbered by  $r$  (see Table 1) in the horizontal axis. The behavior of these descriptors with the number of chlorine atoms is very regular. The average frequency ( $M_{\rho,1}$ ) decreases when the number of chlorine atoms increases. The widths of the spectral density distributions ( $M'_{\rho,2}$ ) also decrease when the number of chlorine atoms increases. The reverse trend is observed for the asymmetry coefficients ( $M''_{\rho,3}$ ) and for the excess ( $M''_{\rho,4}$ ) which increase when the number of chlorine atoms increases. The case of the molecule with eight chlorine atoms is different and the values of its descriptors are the smallest.

Figure 4 shows spectra of 6 compounds: (a) naphthalene  $r = 0$ , (b) 1 chlorine atom  $r = 1$ , (c) 2 chlorine atoms  $r = 3$ , (d) 4 chlorine atoms  $r = 27$ , (e) 7 chlorine atoms  $r = 73$ , (f) 8 chlorine atoms  $r = 75$ .

**Fig. 4** IR spectra: **a**  $r = 0$ ; **b**  $r = 1$ ; **c**  $r = 3$ ; **d**  $r = 27$ ; **e**  $r = 73$ ; **f**  $r = 75$







**Fig. 5** Spectral density distribution moments for 76 compounds calculated using spectra with some errors for  $r = 27, 28, \dots, 75$

In [19] and in [25] one may find, respectively, the assignment of naphthalene vibrational spectrum and an analysis of the IR spectrum of chloronaphthalene. The high frequency oscillations ( $> 3000 \text{ cm}^{-1}$ ) correspond to C–H stretching vibrations. In chlorine substituted compounds these frequencies are gradually replaced by much lower frequency C–Cl stretching modes. This simple fact explains the shift of  $M_{\rho,1}$  towards lower values as a number of chlorine atoms increases. Also out-of-plane (oop) C–H vibrations, located in the naphthalene in the region of  $786\text{--}980 \text{ cm}^{-1}$  (see for example assignments in [26]) gradually disappear while a number of chlorine substituents increases in the molecule. The oop modes for C–Cl fragments have quite low frequencies (around  $100 \text{ cm}^{-1}$ , see Fig. 4) and thus lower  $M_{\rho,1}$  values are expected.

These two facts (replacing high frequency C–H stretching modes by low frequency C–Cl stretching modes and replacing oop modes for C–H by oop modes for C–Cl) are also related to the shift of  $M'_{\rho,2}$  to lower values when the number of chlorine atoms increases. The dispersion of the frequencies becomes smaller for the spectra of compounds with larger number of chlorines.

The unexpected order of  $M''_{\rho,3}$  descriptors [ $M''_{\rho,3}(7 \text{ Cl}) < M''_{\rho,3}(6 \text{ Cl})$ ] may be explained by the fact that in 7 Cl compounds only two non-equivalent systems are expected, namely these in which the only hydrogen atom is located either in position (1) (alpha) or in (2) (beta). Thus, the symmetry of this distribution is relatively high.

On the other hand, in 6 Cl derivatives there are 10 different combinations of 2 hydrogen atoms and these spectra have less symmetric distribution of frequencies.

It appears that the values of moments calculated for the correct spectra, for a given number of chlorine atoms, do not depend on the distribution of these atoms up to several significant figures. However, this is not the case if spectra contain some errors. Though we still do not have a formal proof of this property, it may be used for the detection of errors in the calculated spectra. This may be seen by comparing Figs. 3 and 5. In Fig. 3 all the spectra are correct. In Fig. 5 the correct spectra are for  $s < 4$  i.e.  $r = 0, 1, \dots, 26$ , and they contain errors for  $s \geq 4$  i.e.  $r = 27, 28, \dots, 75$ , where  $s$  is the number of chlorine atoms in the compounds. In particular, the irregular oscillations may be seen for  $M_{\rho,1}''$  and for  $M_{\rho,3}''$  in top panels of Fig. 5.

Summarizing, spectral density distribution moments clearly identify the number of chlorine atoms in the molecules. They can be used as a tool for checking the correctness of the spectra which are used for their creation.

**Acknowledgments** The contributions of DBW and of TP were supported, respectively, by the Alexander von Humboldt Foundation and by the Polish Ministry of Science and Higher Education (grant no. DS/8430-4-0171-1).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

1. M.T. Sisay, L. Peltason, J. Bajorath, *J. Chem. Inf. Model* **49**, 2179 (2009)
2. *Advances in Molecular Similarity*, vol. 2, ed. by R. Carbo-Dorca, P.G. Mezey (JAI Press, Stamford, CN, 1998)
3. M. Johnson, G.M. Maggiora, *Concepts and Applications of Molecular Similarity* (Wiley, New York, 1990), p. 393
4. *Handbook of Chemoinformatics*, ed. by J. Gasteiger, (Wiley-VCH, Weinheim, 2003)
5. *Topological Indices and Related Descriptors in QSAR and QSPR*, ed. by J. Devillers, A.T. Balaban, (Gordon and Breach Science Publishers, The Netherlands, 1999)
6. R.D. Cramer, D.E. Patterson, J.D. Bunce, *J. Am. Chem. Soc.* **110**, 5959 (1988)
7. J.S. Duca, A.J. Hopfinger, *J. Chem. Inf. Comput. Sci.* **41**, 1367 (2001)
8. T. Clark, *J. Comput.-Aided Mol. Design* **24**, 605 (2010)
9. M. Brustle, B. Beck, T. Schindler, W. King, T. Mitchell, T. Clark, *J. Med. Chem.* **45**, 3345 (2002)
10. A.H. Göller, M. Hennemann, J. Keldenich, T. Clark, *J. Chem. Inf. Model* **46**, 648 (2006)
11. T. Clark, K.G. Byler, M.J. de Groot, Proceedings of the International Beilstein Workshop, Bozen, Italy, May 15–19, 2006, in *Molecular Interactions - Bringing Chemistry to Life*, (Logos Verlag, Berlin, 2008) pp. 129–146
12. T. Clark, M. Ford, J. Essex, W. G. Richards, D. W. Ritchie, Proceedings of EuroQSAR, 2004, *QSAR and Molecular Modelling in Rational Design of Bioactive Molecules* ed. by E. Aki, I. Yalcin, (CADDDS in Turkey, Ankara, 2006) pp. 536–537
13. D. Bielińska-Wąż, P. Wąż, S.C. Basak, *Eur. Phys. J. B* **50**, 333 (2006)
14. D. Bielińska-Wąż, P. Wąż, S.C. Basak, *J. Math. Chem.* **42**, 1003 (2007)
15. D. Bielińska-Wąż, P. Wąż, *J. Math. Chem.* **43**, 1287 (2008)
16. D. Bielińska-Wąż, W. Nowak, Ł. Peplowski, P. Wąż, S.C. Basak, R. Natarajan, *J. Math. Chem.* **43**, 1560 (2008)
17. J. Falandysz, *Food Addit. Contam* **20**, 995 (2003)

18. D.M. Hudgins, S.A. Sandford, L.J. Allamandola, *Astronomical Infrared Spectroscopy. Future Observational Directions*, vol. 41. Conference held at the University of Calgary, Alberta, June 16–19, 1992, ed. by S. Kwok (Astronomical Society of the Pacific, San Francisco, CA, 1993), p. 275
19. S.R. Langhoff, *J. Phys. Chem.* **100**, 2819 (1996)
20. P. Wąż, D. Bielińska-Wąż, A. Pleskacz, A. Strobel, *Acta Phys. Pol. B* **39**, 1993 (2008)
21. P. Wąż, D. Bielińska-Wąż, A. Strobel, A. Pleskacz, *Acta Astron.* **60**, 283 (2010)
22. P. Wąż, D. Bielińska-Wąż, *Acta Phys. Pol. A* **116**, 987 (2009)
23. D. Bielińska-Wąż, W. Nowak, P. Wąż, A. Nandy, T. Clark, *Chem. Phys. Lett.* **443**, 408 (2007)
24. M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr, T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Men-  
nucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota,  
R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li,  
J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Strat-  
mann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma,  
G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain,  
O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G.  
Baboul, S. Clifford, J. Ciosłowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L.  
Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W.  
Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople (Gaussian, Inc., Wallingford CT,  
2004). [http://www.gaussian.com/g\\_misc/g03/citation\\_g03.htm](http://www.gaussian.com/g_misc/g03/citation_g03.htm)
25. S.N. Singh, H. Bhatti, R. Singh, *Spectrochim. Acta Part A Mol. Spectrosc.* **985**, 34 (1978)
26. W.X. Huang, J.M. White, *J. Phys. Chem. B* **108**, 5060 (2004)