Int. J. Mol. Sci. **2007**, *8*, 335-345

International Journal of
Molecular Sciences
ISSN 1422-0067

www.mdpi.org/ijms/

Full Research Paper

How Good Can the Characteristic Polynomial Be for Correlations?

Sorana Daniela Bolboaca [1,*] and Lorentz Jantschi [2]

1 "Iuliu Haţieganu" University of Medicine and Pharmacy, 13 Emil Isac, 400023 Cluj-Napoca, Romania
2 Technical University of Cluj-Napoca, 15 Constantin Daicoviciu, 400020 Cluj-Napoca, Romania
E-mail: lori@j.academicdirect.org

[*] Author to whom correspondence should be addressed: E-mail: sorana@j.academicdirect.ro

Received: 14 January 2007 / In Revised Form: 27 March 2007 / Accepted: 12 April 2007 / Published: 30 April 2007

Abstract: The aim of this study was to investigate the characteristic polynomials resulting from the molecular graphs used as molecular descriptors in the characterization of the properties of chemical compounds. A formal calculus method is proposed in order to identify the value of the characteristic polynomial parameters for which the extremum values of the squared correlation coefficient are obtained in univariate regression models. The developed calculation algorithm was applied to a sample of nonane isomers. The obtained results revealed that the proposed method produced an accurate and unique solution for the best relationship between the characteristic polynomial as molecular descriptor and the property of interest.

Keywords: Characteristic polynomial, Graph theory, Structure-Property Relationships, Nonane isomers, Henry's law constant (solubility).

---

1. Introduction

Polynomials derived from molecular graphs and matrixes find applications in chemistry for the construction of structural descriptors and topological indices [1], in QSPR (quantitative structure-

property relationships) and QSAR (quantitative structure-activity relationships) models [2,3]. The characteristic polynomial of a molecular graph is a structural invariant, defined as [4,5]:

$$ChP(G, X) = \det[XI - A(G)] \quad (1)$$

where A(G) is the adjacency matrix (A being a square matrix) of the molecular graph G, and I is the identity matrix.

Crum-Brown and Fraser published in 1898 the observation that the physiological action of ammonium salts is a function of their chemical composition and structure [6]. Since then, many indices have been introduced and used in the characterization of compounds' properties such as the Wiener index [7,8], Hosoya index [9,10], Zagreb index [11,12], Wiener-Hosoya index [13,14], Randić index [15,16], Narumi-Katayama index [17,18], Pogliani index [19], Schultz index [20], Gutman index [21], Harary H index [22], Cluj index [23], Balaban index [24], Xu index [25], and others.

Hosoya first reported the use of the absolute values of the coefficients of the characteristic polynomial of a non-ciclic chemical compound in 1971 [9], known today as the Hosoya index Z. Since then, the analysis of the correlation between Z and many thermodynamic properties has been thoroughly studied [26-31]. However, a polynomial is a more general treatment than an index. The characteristic polynomial is just one polynomial calculated on a molecular structure [5]. The advantage of polynomials is the reduction of degeneration. Our goal was to create a procedure for creating and using a polynomial formula to correlate the structure with a given property through the value of polynomials at a point. This concept generalizes somehow the use of polynomials in regression analysis. Moreover, the desired functionality of our application is to find all singularities of polynomials derivatives, in order to answer our proposed question: How Good Can the Characteristic Polynomial Be for Correlations?

Starting with the characteristic polynomials as molecular descriptors in characterization of structure-property relationships, the aim of the research was to develop a formal calculation algorithm able to identify the value of the characteristic polynomial parameter for which the extremum values of squared correlation coefficients are obtained in univariate regression models.

2. Statement of the Problem and Mathematical Solution

Let's consider a sample of n compounds. The molecule will be abbreviated as c_i where i is an integer and takes values from *1* to n.

The characteristic polynomial can be built and calculated based on the compound's structure by using the following generic functions (where for the simplification all polynomials are of the same degree k):

$$mol_i \quad ChP_i = a_{0i}X^0 + a_{1i}X^1 + a_{2i}X^2 + \ldots + a_{ki}X^k \quad (2)$$

where a_{ki} are coefficients of the characteristic polynomial (a_{0i} = the constant coefficient and a_{ki} = the leading coefficient, k = the degree of polynomial), ChP_i are the characteristic polynomial functions, and X is a generic variable.

Each chemical compound from the sample (c_i) has a molecular structure (S_i) and an associated property of interest (Y_i). These can be written as:

$$mol_i \quad Y_i \quad S_i \quad (3)$$

We have compounds (c_i) with associated property of interest (Y_i), and starting from their structure associated characteristic polynomials (ChP_i):

$$\mathrm{mol_i} \quad \mathrm{Y_i} \quad \mathrm{ChP_i(X)} \tag{4}$$

For characterization of the compounds' property, the abstract function of the characteristic polynomial is not useful; the value associated with the characteristic polynomial function is necessary:

$$\mathrm{mol_i} \quad \mathrm{Y_i} \quad \mathrm{ChP_i}(x) \tag{5}$$

where $ChP_i(x)$ is the value of the characteristic polynomial function associated to *i* molecule.

A problem arises at this point: what is the value of X ($X = x$) for which the correlation between the property of interest and the characteristic polynomial function attain the maximum value?

It is well known that the Pearson product-moment correlation coefficient is the most used correlation coefficient for quantitative variables. In our example this coefficient indicates the strength and direction of the linear relationship between property of interest and characteristic polynomial. Transforming the problem into a formula, the problem becomes:

$$\mathrm{r\left(Y, ChP(X)\right)} = \frac{\mathrm{cov\left(Y, ChP(X)\right)}}{\sigma_{\mathrm{Y}}\sigma_{\mathrm{ChP(X)}}} = \frac{\mathrm{M\left((Y-\mu_Y)(ChP(X)-\mu_{ChP(X)}\right)}}{\sigma_{\mathrm{Y}}\sigma_{\mathrm{ChP(X)}}} \qquad \mathrm{r(Y,ChP(X)) = max} \tag{6}$$

where *cov* is the covariance; σ_y, $\sigma_{ChP(X)}$ are the standard deviation of the property activity (Y) and characteristic polynomial ($ChP(X)$); M is the expected value of the variables Y and $ChP(X)$; and μ_Y, $\mu_{ChP(X)}$ are the variables averages.

The above parameters could be written as: $\mu_Y = M(Y)$, $\sigma_y^2 = M(Y)^2 - M^2(Y)$, and similarly $\mu_{ChP(X)} = M(ChP(X))$, $\sigma_{ChP(X)}^2 = M(ChP(X))^2 - M^2(ChP(X))$. In these conditions, the formula of the correlation coefficient is:

$$\mathrm{r\left(Y, ChP(X)\right)} = \frac{\mathrm{M\left(YChP(X)\right) - M(Y)M\left(ChP(X)\right)}}{\sqrt{\mathrm{M(Y^2) - M^2(Y)}}\sqrt{\mathrm{M(ChP(X)^2) - M^2(ChP(X))}}} \qquad \mathrm{r(Y,\ ChP(X)) = max} \tag{7}$$

To solve the problem it is necessary to find equations of unknown grade in X with real solutions. The formula:

$$\partial \mathrm{r}/\partial \mathrm{X} = 0 \rightarrow \mathrm{x_1, \ldots x_j} \tag{8}$$

where $\partial r/\partial X$ = derivative of $r(Y, ChP(X))$, and j is an integer, gives the solutions for $x_1, \ldots, x_j$.

Note that it is difficult to work with r from Eq. (7); it is much easier to work with its squared value (r^2). Using squared correlation coefficient (r^2) instead of correlation coefficient (r), Eq. (8) becomes:

$$\partial \mathrm{r}^2/\partial \mathrm{X} = 2\mathrm{r}\partial \mathrm{r}/\partial \mathrm{X} = 0 \tag{9}$$

So, the roots $x_1, \ldots, x_j$ of $\partial r(Y, ChP(X))/\partial X = 0$ will be between the roots of $\partial r2(Y, ChP(X))/\partial X = 0$. In any case, not all roots of $r^2 = 0$ (or $r = 0$) are of interest. Eq. (10) will provide all extremum points (Eq. (11)):

$$\partial(\cdot)/\partial \mathrm{X} = 0 \tag{10}$$

$$\partial(\cdot)(\mathrm{Y,ChP(X)})/\partial \mathrm{X}|_{\mathrm{X=x}} = 0 \Leftrightarrow \mathrm{x\ is\ a\ extremum\ point\ of\ (\cdot)} \tag{11}$$

where dot (·) designs any function (such as r, r^2 in our case)

In order to find which among the solutions ($\{x_1, \ldots, x_k\}$) of Eq. (11) are global maxima, the values of all $r(Y,ChP(x_k))$ must be computed and from the obtained values the greatest ones must be selected:

$$x_j \text{ is a maximum (positive or negative)} \Leftrightarrow r(Y,ChP(x_j)) = \max\{r(Y,ChP(x_k))\} \quad (12)$$

Assuming that there is a string of polynomials (as in Eq. (2)) with equal degree *k*:

$$P_j = a_{0j}X^0 + a_{1j}X^1 + a_{2j}X^2 + \ldots + a_{kj}X^k \quad (13)$$

the proposed implementation of the model uses the following elementary mathematical operations:

÷ Multiplication:

$$R = \alpha P_j = \alpha a_{0j}X^0 + \alpha a_{1j}X^1 + \alpha a_{2j}X^2 + \ldots + \alpha a_{kj}X^k \quad (14)$$

÷ Addition:

$$R = P_i + P_j = (a_{0i}+a_{0j})X^0 + (a_{1i}+a_{1j})X^1 + (a_{2i}+a_{2j})X^2 + \ldots + (a_{ki}+a_{kj})X^k \quad (15)$$

÷ Average:

$$R = M(P_i) = M(a_{0j})X^0 + M(a_{1j})X^1 + M(a_{2j})X^2 + \ldots + M(a_{kj})X^k \quad (16)$$

÷ Product:

$$R = P_iP_j = (a_{0i}a_{0j})X^0 + (a_{0i}a_{1j}+a_{1i}a_{0j})X^1 + \ldots + (a_{ki}a_{kj})X^{2k} \quad (17)$$

÷ Derivative:

$$R = P_i' = a_{1j}X^0 + 2a_{2j}X^1 + \ldots + ka_{kj}X^{k-1} \quad (18)$$

In order to solve Eq (9), a derivative of a fraction is also necessary:

$$\text{if } R = (P_i/P_j)' = 0 \text{ then } P_i'P_j - P_iPj' = 0 \quad (19)$$

The proposed calculus could be done with pen and paper, but is time consuming, especially when there are many compounds of interest. Thus, a formal computation method could help to find the exact and unique solution of the best relationship between characteristic polynomial and property of interest.

3. Calculation Algorithm

1. Parse polynomials formulas for all given molecules (ChP_j, $1 \le j \le n$); parse measured data values for given molecules (Y_j, $1 \le j \le n$). Comments:
   a. The polynomials are stored as sums of monomials;
   b. Every monomial is in fact a pair of two values: the power of variable (*X*) and the coefficient;
   c. A measured data value is assigned with a polynomial through *j* value (where *j* is an integer and takes value from *1* to *n*).
2. Search in the polynomial formulas and remove the identical monomials (as in Table 1). Comments:
   a. It is safe to remove the repeated monomials (such for example the X^9 or $-8\cdot X^7$, see Table 1). The calculations made by using Eq. (7) revealed that the values of correlation coefficients are not affected;
   b. It is better to remove the identical monomials in order to reduce the calculation complexity, magnitude of numbers, and errors propagation.
3. Compute the polynomial of squared correlation coefficient formula as pair of two polynomials: numerator and denominator. Comment: The following procedures has been used:
   a. Compute the mean and dispersion of *Y* (as numbers): $mY = M(Y)$; and $d2Y = M(Y^2)-M^2(Y)$;
   b. Compute the average polynomial (as polynomial): $MChP(X) = M(ChP(X))$;

c. Compute the average of *YChP(X)* products (as polynomial): *MYChP(X) = M(YChP(X))*;
d. Construct square polynomials of $ChP_j^2(X)$ and average them (as polynomial): *MChP2(X) = M(ChP²(X))*;
e. Make the product of *MChP* (as polynomial): *M2ChP(X) = MChP(X)·MChP(X)*;
f. Change the sign of *M2ChP* (as polynomial): $M2ChP(X) = (-1)\cdot M^2ChP(X)$;
g. Add M2ChP to *MChP2* (as polynomial): $MChP2(X) = MChP^2(X) + M^2ChP(X)$;
h. Multiply the obtained *MChP2* by *d2Y*: $MChP2(X) = (d2Y)\cdot MChP^2(X)$ // Comment: now the *MChP2(X)* polynomial contains the denominator of r^2;
i. Multiply *MChP* with *(-mY)*: *MChP(X) = (-mY)·MChP(X)*;
j. Add the obtained *MChP(X)* polynomial to the *MYChP(X)* polynomial: *MYChP(X) = MYChP(X) + MChP(X)* // Comment: now the *MYChP(X)* contains the numerator of r;
k. Square the obtained *MYChP(X)* polynomial: *MYChP(X) = MYChP(X)·MYChP(X)* // Comment: now the *MYChP(X)* polynomial contains the numerator of r^2;
l. Return the pair of polynomials (*MYChP(X),MChP2(X)*).

4. Calculate derivative of the numerator of r^2 (as polynomial): *numerator1(X) = ∂numerator(X)/∂X*;
5. Calculate derivative of the denominator of r^2 (as polynomial): *denominator1(X) = ∂denominator(X)/∂X*;
6. Calculate the product between *numerator1(X)* and *denominator(X)* (as polynomial): *product1(X) = numerator1(X)·denominator(X)*;
7. Calculate the product between *numerator(X)* and *denominator1(X)* (as polynomial): *product2(X) = numerator(X)·denominator1(X)*;
8. Change the sign of the *product2(X)*: *product2(X) = (-1)·product2(X)*;
9. Add the *product2(X)* to the *product1(X)* and store the result in the *r2_1_numerator*: *r2_1_numerator(X) = product1(X) + product2(X)*;
10. Factorize *r2_1_numerator(X)* if it is possible (usually is easy to factorize with X if this factor is contain in it, so will factorize on X); let X^p be the factor; delete the factor; thus the *r2_1_numerator* became: $r2\_1\_numerator(X) = r2\_1\_numerator(X)/X^p$;
11. Find roots of *equation r2_1_numerator(X) = 0* and return them as pairs (x_i,ε_i) $1 \le i \le m$ where in fact $r2\_1\_numerator(x_i) = \varepsilon_i$. Comments:
    a. The procedure of finding roots is an approximate one for at least two reasons. First, the *M(·)* operator is used, so the coefficients cannot be integers. Second, even if the *S(·)* operator (sum operator) is used instead of the *M(·)* operator in order to obtain integer coefficients, the degree of the obtained polynomial is too great to apply some nonnumeric methods here (for our example the degree of the obtained polynomial equation was 12);
    b. The returning of the ε_i is used in order to know how close the exact solution is to the result;
    c. The procedure of finding roots is a recursive one and it also calculates and uses all superior derivatives of the polynomial in order to find all real roots of the equation.
12. Use the set of roots $\{x_i\}1 \le i \le m$ and pairs of polynomials (*numerator(X),denominator(X)*) to calculate the value of r^2 in the following points: $\{x_i\}_{1 \le i \le m} \rightarrow \{r^2(x_i)\}_{1 \le i \le m}$
13. Display the results: $\{x_i,\varepsilon_i,r^2(x_i)\}_{1 \le i \le m}$

The above-presented algorithm has been implemented using PHP language (Hypertext Preprocessor). In order to illustrate its effectiveness, the program was run for a sample of nonane isomers, the Henry's law constant (solubility) being the property of interest.

4. Henry's Law Constant of Nonane Isomers: Computational Results and Discussion

Nonane isomers are acyclic saturated hydrocarbon structures with the general chemical formula C_9H_{20}. There are thirty-five compounds in this class: 4-methyloctane (c_1), 3-ethyl-2,3-dimethylpentane (c_2), 3,3-diethylpentane (c_3), 2,2,3,3-tetramethyl-pentane (c_4), 2,3,3,4-tetramethylpentane (c_5), nonane (c_6), 2,3,3-trimethylhexane (c_7), 3,3,4-trimethylhexane (c_8), 3-ethyl-3-methylhexane (c_9), 2,2,3,4-tetramethylpentane (c_{10}), 3,4-dimethylheptane (c_{11}), 2,3,4-trimethylhexane (c_{12}), 3-ethyl-4-methylhexane (c_{13}), 3-ethyl-2,2-dimethylpentane (c_{14}), 3-ethyl-2,4-dimethylpentane (c_{15}), 2,3-dimethylheptane (c_{16}), 3,3-dimethylheptane (c_{17}), 4,4-dimethylheptane (c_{18}), 3-ethylheptane (c_{19}), 4-ethyl-heptane (c_{20}), 2,2,3-trimethylhexane (c_{21}), 2,2,5-trimethylhexane (c_{22}), 2,4,4-trimethylhexane (c_{23}), 3-ethyl-2-methylhexane (c_{24}), 2,2,4,4-tetramethylpentane (c_{25}), 3-methyloctane (c_{26}), 2,5-dimethylheptane (c_{27}), 3,5-dimethylheptane (c_{28}), 2,3,5-trimethylhexane (c_{29}), 2-methyloctane (c_{30}), 2,2-dimethylheptane (c_{31}), 2,4-dimethylheptane (c_{32}), 2,6-dimethylheptane (c_{33}), 2,2,4-trimethyl-hexane (c_{34}), and 4-ethyl-2-methyl-hexane (c_{35}), respectively. The Henry's law constant (solubility of a gas in water) of alkanes expressed as trace gases of potential importance in environmental chemistry was the property of interest. The measured values were taken from a previously reported research [32] (k_H, Table 1) and were given as M/atm unit measurements (M/atm = $[mol_{aq}/dm^3_{aq}]$/atm).

In the fist step of the calculation algorithm, the polynomial formulas for all thirty-five compounds and associated measured Henry's law constants were parsed. After the second step of the computing algorithm, two identical monomials (X^9 and $-8\cdot X^7$) were identified and those monomials were removed from the polynomials (see Table 1, characteristic polynomials after second step - last column).

The polynomial of the squared correlation coefficient resulting from the third step of the calculation algorithm was of the tenth degree:

$$r^2(P(X)) = (X^2\cdot 0.55\ldots - X^4\cdot 0.94\ldots - X^6\cdot 0.39\ldots + X^8\cdot 0.66\ldots + X^{10}\cdot 0.27\ldots)/ \\ /(X^2\cdot 14.19\ldots - X^4\cdot 56.43\ldots + X^6\cdot 100.25\ldots - X^8\cdot 52.97\ldots + X^{10}\cdot 9.93\ldots) \quad (20)$$

The derivative of the r^2 numerator was of the twelfth degree:

$$r^{2\prime}(P(X)) = 0 \\ <=> \\ (-0.84\ldots)X^0 + (5.74\ldots)X^2 + (-10.9\ldots)X^4 + (8.47\ldots)X^6 + (-1.26\ldots)X^8 + (-2.97\ldots)X^{10} + X^{12} = 0 \quad (21)$$

Note that just the first significant digits were displayed in Eqs. (20) and (21) (the "…" sign was written when more digits were available).

The solutions of roots for the squared correlation coefficient obtained by the proposed algorithm for the sample of nonane isomers are presented in Table 2, where the ε_i parameter shows how closely the obtained value is to the exact solution $r^2(x_i)'_X = 0$. Indeed, the *r2_1_numerator*$(x_i) = \varepsilon_i$ was true, where the *r2_1_numerator* was from the eleventh step of the proposed algorithm and represented a part of parameter depicted above.

Table 1. Nonane isomers: Henry's law constant and characteristic polynomials.

Comp. Abbrev.	$k_H(\cdot 10^{-5})$ [M/atm]*	Characteristic polynomial	After second step of calculation algorithm
c_1	10	$X^9 - 8\cdot X^7 + 20\cdot X^5 - 17\cdot X^3 + 3\cdot X$	$20\cdot X^5 - 17\cdot X^3 + 3\cdot X$
c_2	15	$X^9 - 8\cdot X^7 + 17\cdot X^5 - 12\cdot X^3 + 2\cdot X$	$17\cdot X^5 - 12\cdot X^3 + 2\cdot X$
c_3	15	$X^9 - 8\cdot X^7 + 16\cdot X^5 - 8\cdot X^3$	$16\cdot X^5 - 8\cdot X^3$
c_4	16	$X^9 - 8\cdot X^7 + 15\cdot X^5 - 6\cdot X^3$	$15\cdot X^5 - 6\cdot X^3$
c_5	16	$X^9 - 8\cdot X^7 + 18\cdot X^5 - 16\cdot X^3 + 5\cdot X$	$18\cdot X^5 - 16\cdot X^3 + 5\cdot X$
c_6	17	$X^9 - 8\cdot X^7 + 21\cdot X^5 - 20\cdot X^3 + 5\cdot X$	$21\cdot X^5 - 20\cdot X^3 + 5\cdot X$
c_7	17	$X^9 - 8\cdot X^7 + 17\cdot X^5 - 10\cdot X^3$	$17\cdot X^5 - 10\cdot X^3$
c_8	17	$X^9 - 8\cdot X^7 + 17\cdot X^5 - 11\cdot X^3 + 2\cdot X$	$17\cdot X^5 - 11\cdot X^3 + 2\cdot X$
c_9	17	$X^9 - 8\cdot X^7 + 18\cdot X^5 - 14\cdot X^3 + 3\cdot X$	$18\cdot X^5 - 14\cdot X^3 + 3\cdot X$
c_{10}	17	$X^9 - 8\cdot X^7 + 16\cdot X^5 - 6\cdot X^3$	$16\cdot X^5 - 6\cdot X^3$
c_{11}	18	$X^9 - 8\cdot X^7 + 19\cdot X^5 - 15\cdot X^3 + 3\cdot X$	$19\cdot X^5 - 15\cdot X^3 + 3\cdot X$
c_{12}	18	$X^9 - 8\cdot X^7 + 18\cdot X^5 - 12\cdot X^3 + 2\cdot X$	$18\cdot X^5 - 12\cdot X^3 + 2\cdot X$
c_{13}	18	$X^9 - 8\cdot X^7 + 19\cdot X^5 - 16\cdot X^3 + 4\cdot X$	$19\cdot X^5 - 16\cdot X^3 + 4\cdot X$
c_{14}	18	$X^9 - 8\cdot X^7 + 17\cdot X^5 - 10\cdot X^3$	$17\cdot X^5 - 10\cdot X^3$
c_{15}	18	$X^9 - 8\cdot X^7 + 18\cdot X^5 - 12\cdot X^3$	$18\cdot X^5 - 12\cdot X^3$
c_{16}	19	$X^9 - 8\cdot X^7 + 19\cdot X^5 - 14\cdot X^3 + 2\cdot X$	$19\cdot X^5 - 14\cdot X^3 + 2\cdot X$
c_{17}	19	$X^9 - 8\cdot X^7 + 18\cdot X^5 - 12\cdot X^3 + 2\cdot X$	$18\cdot X^5 - 12\cdot X^3 + 2\cdot X$
c_{18}	19	$X^9 - 8\cdot X^7 + 18\cdot X^5 - 12\cdot X^3$	$18\cdot X^5 - 12\cdot X^3$
c_{19}	19	$X^9 - 8\cdot X^7 + 20\cdot X^5 - 18\cdot X^3 + 5\cdot X$	$20\cdot X^5 - 18\cdot X^3 + 5\cdot X$
c_{20}	19	$X^9 - 8\cdot X^7 + 20\cdot X^5 - 18\cdot X^3 + 4\cdot X$	$20\cdot X^5 - 18\cdot X^3 + 4\cdot X$
c_{21}	19	$X^9 - 8\cdot X^7 + 17\cdot X^5 - 9\cdot X^3$	$17\cdot X^5 - 9\cdot X^3$
c_{22}	19	$X^9 - 8\cdot X^7 + 17\cdot X^5 - 6\cdot X^3$	$17\cdot X^5 - 6\cdot X^3$
c_{23}	19	$X^9 - 8\cdot X^7 + 17\cdot X^5 - 8\cdot X^3$	$17\cdot X^5 - 8\cdot X^3$
c_{24}	19	$X^9 - 8\cdot X^7 + 19\cdot X^5 - 15\cdot X^3 + 2\cdot X$	$19\cdot X^5 - 15\cdot X^3 + 2\cdot X$
c_{25}	19	$X^9 - 8\cdot X^7 + 15\cdot X^5$	$15\cdot X^5$
c_{26}	20	$X^9 - 8\cdot X^7 + 20\cdot X^5 - 17\cdot X^3 + 4\cdot X$	$20\cdot X^5 - 17\cdot X^3 + 4\cdot X$
c_{27}	20	$X^9 - 8\cdot X^7 + 19\cdot X^5 - 13\cdot X^3 + 2\cdot X$	$19\cdot X^5 - 13\cdot X^3 + 2\cdot X$
c_{28}	20	$X^9 - 8\cdot X^7 + 19\cdot X^5 - 14\cdot X^3 + 3\cdot X$	$19\cdot X^5 - 14\cdot X^3 + 3\cdot X$
c_{29}	20	$X^9 - 8\cdot X^7 + 18\cdot X^5 - 10\cdot X^3$	$18\cdot X^5 - 10\cdot X^3$
c_{30}	21	$X^9 - 8\cdot X^7 + 20\cdot X^5 - 16\cdot X^3 + 2\cdot X$	$20\cdot X^5 - 16\cdot X^3 + 2\cdot X$
c_{31}	21	$X^9 - 8\cdot X^7 + 18\cdot X^5 - 10\cdot X^3$	$18\cdot X^5 - 10\cdot X^3$
c_{32}	21	$X^9 - 8\cdot X^7 + 19\cdot X^5 - 13\cdot X^3$	$19\cdot X^5 - 13\cdot X^3$
c_{33}	21	$X^9 - 8\cdot X^7 + 19\cdot X^5 - 12\cdot X^3$	$19\cdot X^5 - 12\cdot X^3$
c_{34}	21	$X^9 - 8\cdot X^7 + 17\cdot X^5 - 7\cdot X^3$	$17\cdot X^5 - 7\cdot X^3$
c_{35}	21	$X^9 - 8\cdot X^7 + 19\cdot X^5 - 14\cdot X^3 + 2\cdot X$	$19\cdot X^5 - 14\cdot X^3 + 2\cdot X$

* $M/atm = (mol_{aq}/dm^3_{aq})/atm$

Table 2. Algorithm of calculation: solutions for nonane isomers.

Solution	x_i	ε_i	$r^2(x_i)$
1.1	- 1.656…	$- 5.5\ldots\cdot10^{-11}$	0.296…
2.1	- 0.856…	$1.1\ldots\cdot10^{-13}$	0
3.1	- 0.481…	$2.7\ldots\cdot10^{-13}$	0.055…
3.2	0.481…	$2.7\ldots\cdot10^{-13}$	0.055…
2.2	0.856…	$1.1\ldots\cdot10^{-13}$	0
1.2	1.656…	$- 5.5\ldots\cdot10^{-11}$	0.296…

x_i = root; $r^2(x_i)$ = squared correlation coefficient; ε_i = numerical error;

… = for all numbers only first significant digits were presented

As it can be observed from Table 2, the proposed algorithm obtained pairs of roots (as negative and positive values: 1.1 - 1.2, 2.1 - 2.2, and 3.1 - 3.2, see the values from the x_i column). The values of squared correlation coefficients are local extremum values (maximum and/or minimum values): one negative (for pair of roots of ± 0.856…) and two positive (one minimum for the ± 0.481… pair of roots and one maximum for the ± 1.656… pair of roots). These are the expected results taking into account that the *r2_1_numerator(X)* is a polynomial pair of *X*.

Analyzing the results presented in Table 2 it can be observed that, for the identified roots, the numerical errors of the models were in all cases less than 0.0001. These results sustain the power of the model to identify the imposed solutions. Looking at the values of the obtained squared correlation coefficients it can be observed that the proposed method identified one maximum value (for roots ± 1.656…) and two minimum values (± 0.856, and ± 0.481) (note that these are local extremum values). Regarding the maximum value of the squared correlation coefficient, it can be observed that is 0.296 and, from the statistical point of view, revealed a week linear relationship between the characteristic polynomial and Henry's law constant for the studied alkanes. It must be noted that the aim of the paper was not to obtained a significant correlation coefficient; it was to develop and implement a formal algorithm able to identify the characteristic polynomial parameter for which the extremum values (as maximum and minimum values) for the correlation coefficient are obtained in univariate regression models, this aim being accomplished.

Regarding the proposed method one question can arise: why use the proposed method when the Hosoya Z index [9] can be used in QSPR without using a computer? First, the use of characteristic polynomials instead of the Z index reduces the degeneration. Second, the proposed model is able to find all singularities of polynomial derivatives. Last in sequence but not least in importance, the proposed computer based method is able to work with small as well as with large sample sizes without any involvement of human time or abilities, eliminating any human errors.

It is well known that the squared correlation coefficients increase with the number of variables used by a linear regression model [33]. Starting from this hypothesis it will be interesting to analyze the applicability of the proposed model to multivariate regression models. The next plan of our research refers the implementation of a similar computational algorithm for multivariate models when characteristic polynomials are use as molecular descriptors. Another question that needs to be

answered refers the usefulness of the method for characterization of relationships between compound's activity and structure, an approach that will be investigated in future research.

5. Concluding Remarks

The proposed calculation algorithm is able to obtain unique and reproducible solutions. The solutions are unique, meaning that for a sample of compounds with a property of interest the maximum value of the squared correlation coefficient between property and characteristic polynomials is always given by a single pair of roots. The computation algorithm can be applicable on any class of compounds when the characteristic polynomials are used as descriptors in analysis of the relationship between compounds' structure and their properties.

Acknowledgement

The research was partly supported by UEFISCSU Romania through project ET46/2006.

References

1. Balaban, A.T.; Ivanciuc, O. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J. and Balaban, A.T., Eds.; Gordon & Breach: Amsterdam, **1999**; Chapter 2, 21-57.
2. Bonchev, D. *Information Theoretic Indices for Characterization of Chemical Structure*; Research Studies Press – Wiley: Chichester, UK, **1983**.
3. Kier, L.B.; Hall, L.H. *Molecular Connectivity in Structure-Activity Analysis*; Research Studies Press: Letchworth, **1986**.
4. Schwenk, A.J. *Computing the characteristic polynomial of a graph, Graphs and Combinatorics*; Bari, R.; Harary, F., Eds.; Springer, Berlin, **1974**; pp. 153-172.
5. Diudea, M.V.; Gutman, I.; Jäntschi, L. *Molecular Topology*; Nova Science: Huntington, New York, **2002**; Chapter 3, pp. 53-100.
6. Crum-Brown, A.; Fraser, T.R. On the connection between chemical constitution and physiological action. Part 1. On the physiological action of the salts of the ammonium bases, derived from Strychnia, Brucia, Thebia, Codeia, Morphia, and Nicotia. *T. Roy. Soc. Edin.* **1868**, *25*, 151-203.
7. Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69(1)*, 17-20.
8. Wang, H.; Yu, G. All but 49 numbers are wiener indices of trees. *Acta Appl. Math.* **2006**, *92(1)*, 15-20.
9. Hosoya, H. Topological index, a newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332-2339.
10. Hosoya, H.; Kawasaki, K.; Mizutani, K. Topological index and thermodynamic properties. I. Empirical rules on the boiling points of saturated hydrocarbons. *Bull. Chem. Soc. Jpn.* **1972**, *45*, 3415-3421.
11. Gutman, I.; Trinajstić, N. Graph theory and molecular orbitals. Total φ-electron energy of alternant hydrocarbons. *Chem. Phys. Lett.* **1972**, *17*, 535-538.

12. Nikolić, S.; Kovacević, G.; Milicević, A.; Trinajstić, N. The Zagreb indices 30 years after. *Croat. Chem. Acta* **2003**, *76*, 113-124.
13. Randić, M. Wiener-Hosoya index - A novel graph theoretical molecular descriptor. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 373-377.
14. Westerberg, T.M.; Dawson, K.J.; McLaughlin, K.W. The Hosoya Index, Lucas Numbers, and QSPR. *Endeavor* **2005**, *1*, 1-15.
15. Randić , M. On characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609-6615.
16. Taherpour, A.; Shafiei, F. The structural relationship between Randić indices, adjacency matrixes, distance matrixes and maximum wave length of linear simple conjugated polyene compounds. *J. Mol. Struct. THEOCHEM* **2005**, *726*, 183-188.
17. Narumi, H. New Topological Indices for Finite and Infinite Systems. *MATCH Commun. Math. Comput. Chem.* **1987**, *22*, 195-207.
18. Tomovic, Z.; Gutman, I. Narumi-Katayama index of phenylenes. *J. Serb. Chem. Soc.* **2001**, *66*, 243-247.
19. Pogliani, L. Modeling with Special Descriptors Derived from a Medium-Sized Set of Connectivity Indices. *J. Phys. Chem.* **1996**, *100*, 18065 -18077.
20. Schuttz, H.P. Topological organic chemistry. 1. Graph theory and topological indices of alkanes. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 227-223.
21. Gutman, I.J. Selected properties of the Schultz molecular topological index. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1037-1039.
22. Plavšić, D.; Nikolić, S.; Trinajstić, N.; Mihalić, Z. On the Harary index for the characterization of chemical graphs. *J. Math. Chem.* **1993**, *12*, 235-250.
23. Jäntschi, L.; Katona, G.; Diudea, M.V. Modeling Molecular Properties by Cluj Indices. *MATCH Commun. Math. Comput. Chem.* **2000**, *41*, 151-188.
24. Balaban, A.T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1982**, *89*, 399-404.
25. Ren, B.A. New Topological Index for QSPR of Alkanes. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 139-143.
26. Hosoya, H.; Gotoh, M.; Murakami, M.; Ikeda, S. Topological Index and Thermodynamic Properties. 5. How Can We Explain the Topological Dependency of Thermodynamic Properties of Alkanes with the Topology of Graphs? *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 192-196.
27. Gao Y.D.; Hosoya, H. Topological Index and Thermodynamic Properties. IV. Size Dependency of the Structure-Activity Correlation of Alkanes. *Bull. Chem. Soc. Jpn.* **1988**, *61*, 3093-3102.
28. Narumi, H.; Hosoya, H. Topological Index and Thermodynamic Properties. III. Classification of Various Topological Aspects of Properties of Acyclic Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1985**, *58*, 1778-1786.
29. Narumi H.; Hosoya, H. Topological Index and Thermodynamic Properties. II. Analysis of the Topological Factors on the Absolute Entropy of Acyclic Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1980**, *53*, 1228-1237.

30. Hosoya, H.; Kawasaki, K.; Mizutani, K. Topological Index and Thermodynamic Properties. I. Empirical Rules on the Boiling Point of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1972**, *45*, 3415-3421.
31. Mekenyan, O.; Bonchev, D.; Trinajstić, N. Chemical graph theory: Modeling the thermodynamic properties of molecules. *Int. J. Quantum Chem.* **2004**, *18*, 369-380.
32. Yaws, C.L.; Yang, H.C. In *Thermodynamic and Physical Property Data*; Yaws, C. L. Ed.; Gulf Publishing Company: Houston, TX, USA, **1992**; pp. 181-206.
33. Hawkins, D.M. The Problem of Overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1-12.