Global Journals LaTeX JournalKaleidoscope™
Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

Modeling Retention Indices of a Series Components Food and Pollutants of the Environment: Methods; OLS, LAD

Fatiha Mebarki[1], Khadidja Amirat[2], Salima Ali Mokhnache[3], Djelloul Messadi[4] and Khadidja Amirat[5]

[1] University Badji Mokhtar Annaba

Received: 16 December 2015 Accepted: 1 January 2016 Published: 15 January 2016

Abstract

The gas chromatographic retention indices for (89 pyrazines of test and 25 of validation) on O V-101 and Carbowax -20M are successfuty modeled with the ald of a computer and the Software system. Structural descriptors are calculated and multiple linear regression analysis are used to generate model equations relating structural features to observed retention characteristics then was treated with two methods. The detection of influential observations for the standard least squares regression model is a problem which has been extensively studied. LAD regression diagnostics offers alternative dicapproaches whose main feature is the robustness. Here a nonparametric method for detecting influential observations is presented and compared with other classical diagnostics methods. Comparisons are between models generated for the two stationary was carried out with two methods, and descriptors that may encode differences in solute interactions with stationary phases of differing polarity are discussed and validated results in the state approached by the tests statistics: Test of Anderson-Darling, shapiro-wilk, Agostino, Jarque-Bera and the confidence interval thanks to the concept of robustness to check if the distribution of the errors is really approximate.

Index terms— LAD Regression, Robustness, Outliers, Leverage points, tests statistics.

1 Introduction

ome compose food are volatile heterocyclic which are found in a natural way in our environment and the attraction which the men test for the flavours is ever contradicted during centuries and which have an interest in multiple fields, in particular in the food like flavour. Their presence in food results mainly, of process requiring a stage of cooking (partial or supplements), Egyptian civilization already used them for the kitchen.

In the evaluation of the environmental risks, information on the fate in the environment, the properties, the behavior and the toxicity of a chemical substance is fundamental need.

The volatile heterocyclic also constitutes a significant family of odorous molecules, particularly interesting in the field of the chemistry of the flavours. They represent more than one quarter of the 5000 volatile compounds insulated and characterized to date in our food.

Pyrazines are heterocyclic very present in our food. More than 80 derived from pyrazines were identified in a great number of cooked food, like the bread, the meat, the torrefied coffee, the cocoa or the hazel nuts; they are very powerful aromatizing compounds. Mihara and Enomoto (1985), described a relation structure/retention for a unit of substituted pyrazines for which the increments of indices relating to various substituents on the cycle were given for a small series of substituents present. The method was then extended to integrate others substituents, by adding a term which takes account of the position on the cycle of a substituent compared to the others (Mihara & Masuda, 1987). In a similar approach, Masuda and Mihara (1986) describe the use of indices of connectivity modified to calculate in advance the indices of retention of a series of substituted pyrazines. The methods lead

to good results, in so far as the increments of indices determined in experiments available for the unknown compounds are implied, which constitutes their principal defect. Stanton and Jurs (1989), used methodology QSRR to develop models connecting the structural characteristics of 107 variously substituted pyrazines, with their indices of retention obtained on two columns of very different polarities (OV-101 and Carbowax-20M). The equations were calculated using the multilinear regression, the choice of the explanatory variables (topological, electronic and physical properties) being realized by progressive elimination **??**Swall & Jurs, 1983), among the 85 individual molecular descriptors obtained for each whole molecule. The indices of retention (IR) obtained on each column were treated separately, while drawing from the same sets of descriptors. The models calculated with 6 explanatory variables provide high standards errors (S = 23 units of index -u.i. -on OV-101 and S = 36.33 u.i. out of Carbowax -20 M) which do not predict good predictive capacities for these models, and which let suppose nonlinear relations between descriptors and property (IR) studied.

The objective of this work aims at using methodology QSRR, the approach Method LAD /Least square (LAD/OLS), to model the indices of retention of (114) pyrazines reported from Davit T. Stanton and Peter C.Jurs (1989) and reported from Mihara and Enomoto pyrazines of test and 25 of validation) on O V-101 and Carbowax -20M are successfuty modeled with the ald of a computer and the Software system. Structural descriptors are calculated and multiple linear regression analysis are used to generate model equations relating structural features to observed retention characteristics then was treated with two methods. The detection of influential observations for the standard least squares regression model is a problem which has been extensively studied. LAD regression diagnostics offers alternative dicapproaches whose main feature is the robustness. Here a nonparametric method for detecting influential observations is presented and compared with other classical diagnostics methods. Comparisons are between models generated for the two stationary was carried out with two methods, and descriptors that may encode differences in solute interactions with stationary phases of differing polarity are discussed and validated results in the state approached by the tests statistics: Test of Anderson-Darling, shapiro-wilk, Agostino, Jarque-Bera and the confidence interval thanks to the concept of robustness to check if the distribution of the errors is really approximate. **??**1985), the molecular descriptors being only calculated starting from the chemical structure of the compounds.

The linear statistical model for fixed purposes will be examined by two robust methods for the evaluation of the parameters of regression starting from estimates of the robust coefficients of regression most popular by the appendices. We based ourselves on the comparison between the two methods, the applicability (DA) will be discussed using the diagram of Williams who represents the residues of prediction standardized according to the values of the levers (hi) **??**Eriksson et al..2003; **??**ropsha et al.2003). We present the tests statistics and graph of compatibility at the normal law for validated the results of the state approached between the two methods for a risk ?= 5%.

2 II.

3 Methodology i. Descriptor Generation

One used the molecular software of modeling Hyperchem 6.03, for to represent the molecules, then using semi-empirical method AM1 **??**Dewar et al.,. 1985;. Holder 1998) to obtain the final geometries. It is established (Levine, 2000) that this Method gives good results when one treats small molecules (of less than one hundred atoms), like those considered in this work.

The optimized geometries were transferred in the software dragon from data-processing software version 5. **??**[19], for the calculation of 1320 descriptors while operating on 89 pyrazines of test; subsets of descriptors were chosen by genetic algorithm, these descriptors can be separate in four categories: topological descriptors of The topological, geometrical, physical, and electronic accounts of way and molecular indices of connectivity included. The geometrical descriptors included sectors of shade, the length with the reports/ratios of width, volumes of van der Waals, the surface, and principal moments of inertia. The calculated descriptors of physical property included the molecular refringency of polarizability and molar. The electronic descriptors included most positive and most negative described by **??**aliszan. By employing the software Mobydigs (Todeschini et al., 2009) [21] and by maximizing the coefficient of prédiction Q 2 and minimal R 2 of S (the error).

4 ii. Regression Analysis

The analysis of the multiple linear regressions was carried out with two methods by software Matlab (R2009a) for (LAD) and Minitab 16 for (OLS).

One considers the multiple model of regression given by [9]:?? ?? = ?? 0 + ? ?? ?? ???1 ?? =1 ?? ???? + ?? ??**(1)**

The detection of meaningless statements and 'with action leverage according to the method of least squares is a problem which' was largely studied. The diagnosis by the regression LAD offers alternative approaches whose principal characteristic is the robustness. In our study a non-parametric method to detect the meaningless statements and the point's lever was applied and compared with the traditional method of diagnosis (least squares) [9].

iii. Method of least squares OLS This one was carried out with the software Minitab 16 [33], method MLR applied to the multiple regression consists in defining the ? estimate which minimizes **??**[9, 17, 18]:? ???? 2 = ? (yi-?0-? ????????) 2**(2)**

iv. Least Absolute Deviations (LAD)

The analysis of linear regression multiple was carried out with the software Matlab (R2009a) [31], by using the method of the least variations in absolute value, said method LAD (Least Absolute Deviations), is one of the principal alternatives to the method of least squares when it is a question of estimating the parameters of a model of regression, which minimizes the absolute values and not the values with the square of the term of erreur. La method stable-lad applied to the multiple regression consists in defining the ? estimates which minimize **??**9, 17, the 18]:?|????| = ? |yi-?0-? ????????**(3)**

III.

5 The Data Set

One uses the molecular software Hyperchem The retention data for the114 compounds chromatographed on the OV-101 and CRW-20M stationary phases were taken from (113 taken from Davit T. Stanton and Peter C. Jurs (1) and 1 compound (2-VinylPyrazine) taken from Mihara and Enomoto [29]) and are listed in table 1.

6 IV.

7 Results and Discussion

An ideal model is one that has a high R value, allow standard error, and the fewest independent variables [1,9]. The best models found has 3 descriptors for each stationary phase by using the software Moby Digs [21] are given below. The criterion for identifying a compound as an outlier was that compound being flagged by three or more of six standard statistical tests used to detect outliers in regression analysis. These tests were (1) residual, (2) standardized residual, (3) Studentized residual, (4) leverage, (5) DFFITS, (6) Cook's distance. The residual is the difference between the actual value and the value predicted by the regression equation. The standardized residual is the residual divided by the standard deviation of the regression equation. The Studentized residual is the residual of a prediction divided by its own standard deviation.

Leverage allows for the determination of the influence of a point in determining the regression equation. DFFITS describes the difference in the fit of the equation caused by removal of a given observation, and Cook's distance describes the change in a model coefficient by the removal of a given point. The definition of each descriptor is given table 2: The coefficient of multiple determinations (R 2) indicates the amount of variance in the data set accounted for by the model. The standard error of the regression coefficient is given in each case, and n indicates the number of molecules involved in the regression analysis procedure [1,9].

a) The best models IR(OV-101): (MPC03, X1sol, GATS5e, AEigp, L3e,Qpos); -S=20.892, R 2 =99.30, n=89 compounds. IR(RWC): Se, Mp, X1sol, DP01, Mor06v, Tm; S=22.64, R 2 =99.22, n=89 compounds.

Indeed Figure **??** reproduced the distributions of the standard residues di (ordinary residue report /root of the average square of the variations) according to the adjusted values, which seem random (without particular tendencies).That shows the constancy of variances ? 2 , it be-with saying their independence of the regresses and the adjusted dependent variable.

The quasi-linearity (R = 0, 9951; OV-101 -R = 0, 9835; Carbowax-20M -critic = 0, 96048) of the diagram of the normal scores (Figure **??**) is an index of normality. Values of the statistics of Durbin-Watson (Durbin, & Watson, 1951), [d= 1,33535; OV-101/D = 1,66161; Carbowax-20M] are the greater than higher values given by the tables, respectively for 3 regresses, and any reasonable risk ?, which establishes each time the independence of the residues. The diagnostic statistics joined together in Table **??** make it possible to make comparisons and to draw several conclusions [21]. Values of R 2 and of R 2 (adj) show, each time, quality of adjustment, whereas the very weak differences between R 2 and Q 2 inform about the robustness of the models which are, moreover, very highly significant (high values of the statistics F of Fisher).

Moreover, the similarity of SDEP and SDEC mean that the internal capacities of prediction models are not too dissimilar their capacities of adjustment.

The validation by bootstrap (Q BOOT) confirms all at the same time the capacity of internal prediction and the stability of the models.

8 b) Robust Regression

Any robust method must be reasonably effective once compared to the estimators of least squares; if the fundamental distribution of the errors is normal and primarily more effective independent than the estimators of least squares, when there are peripheral observations. There are various robust methods for the evaluation the parameters of regression. The principal goal of this section is the method LAD (nap of the absolute values of the errors) whose coefficient of regression qualifies the robustness among the additional data [16].

i

9 . Comparison Robust Regression of OLS and LAD

More particularly we will test 2 methods of estimate for the vector of the Parameters ((?? 0 * ,?? 1 * , ? , ?? ?? *):

-Method of least squares ordinary, more known and the most used.

-The method LAD (Sum of the absolute values of the errors.)

The large advantage of the method LAD is his robustness, i.e. that the estimators are not impact by the extreme values, (they are known as "robust"). It is thus particularly interesting to use the method LAD if one is in the presence of aberrant values in comparison with method OLS [8].

10 ii. Comparison of hyperplanes of regression

Each equation on each column check the assumptions on the same linear statistical model for Fixes purposes for each method in comparison with the hyperplane calculated by LAD compared to the hyperplane calculated by the method of least squares.

It is noticed that ??the calculated OLS are not very different for the regression with ?? the LAD on the two columns, except, ??1 the calculated OLS is almost the same ones as for the regression with ??1 the LAD on column CRW and ??4 the calculated OLS is almost the same ones as for the regression with ??4 the LAD on column OV-101.

It is thus relevant to remake a checking of the presences of aberrant values by using the following stage (figure **??**):

The hyperplane of regression can radically change, with the change of the coefficients of the hyperplane.

iii

11 . Graphical Comparisons of Alternative Regression Models

The field of application was discussed using the diagram of Williams. The analysis of the residues shows that the observations (82,25) residues raised but it (48) point influence in the two estimates and the observation (12) point influence with the LAD estimate and lever by least square also observation 4 residue raised with OLS and not lever with LAD in the whole of validation on column OV -101 and on column CRW -20M the observations (45) not influence in the two estimates and observation 16 point influence in the two estimates in the whole of validation.

After elimination of the aberrant points collective between the two methods and after the secondary treatment one has the observation (12) point influence and the observations (1,24) residues raised in the two estimates but it (25) observation 4 residue raised with OLS and not lever with LAD also the observation 4 residue raised in the whole of validation in the two estimates on column OV -101 and on column CRW -20M the observations (45) not influence in the two estimates and observation 16 point influence in the two estimates in the whole of validation and on column CRW -20M the observations (24 **??**5 35) residues raised but it (84)point influence in the two estimates and observation 8 point influence in the two estimates in the whole of validation.

Thus finally the models in which the meaningless statements were removed become after elimination of the aberrant points collective [OV-101: test - (1,12,24), validation (4), CRW-20M: test - **??**24, 25, 35 84), validation (8)

It is noticed besides that ?? the OLS calculate more to approach which for the regression with ?? the LAD on the two columns into precise (??1 , ??3 ?????? ??4) the OLS calculate are almost the same ones as for the regression with (??1 , ??3 ?????? ??4) the LAD and on the same order with (??0 , ??5 ?????? ??6) on OV 101 and ??1the OLS calculate are almost the same ones as for the regression with ??1 the LAD on CRW -20M.

The analysis of the residues shows that in this case All the point of lad method between (-2, 2), but it the analysis of the residues of OLS method shows that the observations [OV-101: test -(6,42), CRW-20M: test -(22, 24, 67 ,78), validation **??**7 ,13,14)] the LAD estimate given good result On the other hand estimate OLS figure (4): iv. Graphical Comparisons of Alternative Regression Models 0,9 0,8 0,7 0,6 0,5 0,4 0,3 0,2 0,1 0,0 Column OV -101 Columns RW -20M Method LAD and OLS (test, validation) Fig. 4 : Diagram of Williams of the residues of prediction standardized according to the lever Lastly, it is noted that LAD is a robust estimator but loses stability in the presence of points aberrant.

We note however the observation that the estimate the least square is near to the LAD estimate to which removed the aberrant values.

To conform the approach between the two methods and to deduce the robust method between them, There is a package of tests of normality (of the standard errors or residues?) indeed, thanks to the concept of robustness, we can used simple techniques (descriptive e.g. statistics, technical graphs) to check if the distribution of the data is really approximate.

Any test is associated a ? risk known as of first species years works us, we will adopt it risk ? = 5%.

12 c) Comparisons of the Tests of normality of the errors between the method LAD and OLS in the approached state

The software Minitab 16 carries out automatically the estimate of the two principal parameters of the normal law (? the Mean (OV-101:0, CRW-20M:0), ? the variation-type(OV-101:13.26, CRW-20M:18.53) for OLS one applying the same principle with the method LAD but one used (it median (OV-101:-0.96, CRW-20M:0.01) ? variation-type (OV-101:13.84, CRW-20M:18.66) and with the number principal in the state approached to the two columns n=32.

13 i. Test statistical a. Test of Anderson-Darling

In our work, one finds us that AD [OV -101: (lad) = 0.250 with value of p>0.250, (OLS) = p=0.938 with value of p = 0.747, n=82]-RCW-20M: (lad) = 0.547 with value of p > 0.250, (OLS) = 0.165 with value of p=0.572 n=84] <AD critique=0.752 with p> 0.1. To 5%, the assumption of normality is compatible with the method LAD and OLS [33,34,35].

14 b. test of Shapiro-Wilk

It is particularly powerful for small manpower (n<50) for this that one using for valid the results of the validation.

For a risk ? = 0. 05, the critical points read in the table of Shapiro-Wilk for n = 23 is W crit = 0. 914 and for n=24 and W crit = 0. 916. In our works, on (OV) [W LAD =0.9969, W MLR = 0.9877, n=24] and on CRW [W LAD = 0, 0.997, W MLR = 0,9227, n=23] W> W crit , with the risk of 5%, the assumption of normality compatible with us is given (normal law) [34,35].

15 c. Test of D'Agostino

For ? = 0.05, the threshold critic is ?2 0:95(2) = 5.99.In our works, on (OV) [:(W LAD = 0,0072 with value of p = 0,99, W OLS = 0.042 with value of p = 0.97, n=82),:] and on CRW [(W LAD = 0,1202 with value of p = 0.94, W OLS = 0,00116 with value of p = 0.99, n=84),] W <Wcrit, with p > 0.1 with the risk of 5%, the assumption of normality compatible with us is given (normal law) [33,34,35].

16 d. Test of Jarque-Bera

As the Test of Agostino It becomes particularly effective starting from N>20 for this that one using for valid the results.

For ? = 0.05, the critical point is ?2 0:95(2) = 5.99. In our works, on (OV)[(W LAD = 0,0971with value of p = 0.95, W OLS = 0.0949 with value of p = 0.95, n=82),] and on CRW [(W LAD = 0.1059 with value of p = 0.94, W OLS = 0,0979 with value of p = 0.95, n=84),] W <W crit (is largely lower than 5.99) with p > 0.1 than the risk of 5%, the assumption of normality compatible with us is given (normal law). [33,34,35] Completely all the statistical tests is accepted the data of the state approached between the two methods especially the test of Shapiro-Wilk the value of the method LAD closer to method OLS and the other tests the values of the method LAD is higher has the method MLR which explains than give them method LAD is effective and robust para for give method OLS.

Completely all the statistical tests is accepted the data of the state approached between the two methods especially the test of Shapiro-Wilk the value of the method LAD closer to method MLR and the other tests the values of the method LAD is higher has the method OLS which explains than give them method LAD is effective and robust para for give method OLS.

e.

17 Interval of confidence

The confidence interval and the risk ?? = 0.05 constitute a complementary approach thus (an approach of estimate) the most used confidence interval is the confidence interval has 100 (1 -?) = 95 %.

The Column OV-101: LAD :(-28.11, 26.17), OLS (-25.9, 25.99).

The Column CRW-20M: LAD (-36.56, 36.58), OLS **??**-36.34, 36.34).

These result is formed L approximate of two method.

You can be 95% confident that the 50th percentile for the population is between OV-101 (LAD:-3.96 and 2.027,-OLS:-2.87 and 2.87, CRW-20M (LAD:-3.98 and 4.00, OLS:-3.96 and 3.96) [33,34,35].

V.

18 Conclusion

The modeling of the indices of retention of 114 pyrazines (89 tests and 25 validations) eluted out of two columns various OV -101 and CRW-20M by two methods LAD and OLS are based on the following comparisons:

a) The comparison of the equations of the hyperplanes L equations of OLS is closer to LAD after elimination of the aberrant points for the ?2 (LAD) ??2(OLS) and the other coefficient remaining with the same order for column OV-101 Pour the column Crw-20m the ?1 (LAD) ??1(OLS) and the other coefficient remaining with the

same order after the secondary treatments for the checking of the presence of aberrant values (82, 48, 26, 25, 24,12, 1) on column OV -101 and item **??**45, **??**2,35,24,25) for the column CRW-20M, and to be able to compare them By employing the following stage.

19 b) Graphic comparison: The applicability was discussed using the diagram of Williams in dependence

Lastly, it is noted that LAD is a robust estimator but loses his stability in the presence of aberrant points.

Used test of normality's of the errors by statistical test. One applied compatibility with the normal law, but to differing degrees using p-been worth. One notes that the touts test to accept the assumption of normality is that of Anderson-Darling, the test of Shapiro-Wilk His power is recognized in the literature.

Lastly, the tests of Agostino and Jarque-Bera, based on the coefficients of asymmetry and flatness accepts readily the assumption of normality with one p-been worth sup 0.1 on the columns, Too one confirmed approached graphically by histogram of frequency in finished by the confidence interval.

It general this study is shown that results by the two estimates theoretical (equation) and graph give good results expressed by the models. [1] [2]

5

Figure 1: 54 5 -

[1] Modeling Retention Indices of a Series Components Food and Pollutants of the Environment: Methods; OLS, LAD © 2016 Global Journals Inc. (US)

[2] Modeling Retention Indices of a Series Components Food and Pollutants of the Environment: Methods; OLS, LAD

1

n°	Compounds	ov-101	Compounds	IR(cw)
1	Pyrazine	710	Pyrazine	
2 3	Methylpyrazine 2,3-	801	Methylpyrazine 2,3-	Year
4	dimethylpyrazine 2,5-	897	dimethylpyrazine 2,5-	2016
	dimethylpyrazine	889	dimethylpyrazine	
5	2,6-dimethylpyrazine	889	2,6-dimethylpyrazine	
6	Trimethylpyrazine	981	Trimethylpyrazine	19
7 8	Trimethylpyrazine Ethylpyrazine	1067	Trimethylpyrazine Ethylpyrazine	IR(cw)
9	2-ethyl-5-methylpyrazine 2-ethyl-	894	2-ethyl-5-methylpyrazine 2-ethyl-	Volume
10	6-methylpyrazine 2,5-dimethyl-	980	6-methylpyrazine 2,5-dimethyl-	XVI
11	3-ethylpyrazine Compounds	977	3-ethylpyrazine Compounds	Is-
n°	2,6-dimethyl-6-ethylpyrazine	1059	2,6-dimethyl-6-ethylpyrazine	sue
12	2,3-dimethyl-5-ethylpyrazine	ov-	2,3-dimethyl-5-ethylpyrazine	I
13	2,3-diethylpyrazine 2,3-diethyl-5-	101	2,3-diethylpyrazine 2,3-diethyl-5-	Ver-
14	methylpyrazine Propylpyrazine	1064	methylpyrazine Propylpyrazine	sion
15		1066		I
16		1065		
		1137		
		986		
17	2-methyl-3-propylpyrazine 2,3-	1072	2-methyl-3-propylpyrazine 2,3-	(
18	dimethyl-5-propylpyrazine	1154	dimethyl-5-propylpyrazine	B
				)
19	2,5-dimethyl-3-propylpyrazine	1142	2,5-dimethyl-3-propylpyrazine	Global
20	2,6-methyl-3-propylpyrazine	1151	2,6-methyl-3-propylpyrazine	Jour-
21	Isopropylpyrazine 2,3-dimethyl-5-	949	Isopropylpyrazine 2,3-dimethyl-5-	nal
22	isopropylpyrazine Butylpyrazine	1112	isopropylpyrazine Butylpyrazine	of
23	2-butyl-3-methylpyrazine 3-	1088	2-butyl-3-methylpyrazine 3-	Hu-
24	butyl-3,5-dimethylpyrazine	1121	butyl-3,5-dimethylpyrazine	man
25	3-butyl-3,6-dimethylpyrazine	1184	3-butyl-3,6-dimethylpyrazine	So-
26	5-butyl-2,3-dimethylpyrazine	1196	5-butyl-2,3-dimethylpyrazine	cial
27	Isobutylpyrazine 2,3-dimethyl-	1254	Isobutylpyrazine 2,3-dimethyl-	Sci-
28	5-isobutylpyrazine 2-isobutyl-	1043	5-isobutylpyrazine 2-isobutyl-	ence
29	3,5,6-trimethylpyrazine sec-	1200	3,5,6-trimethylpyrazine sec-	-
30	butylpyrazine 5-sec-butyl-2,3-	1263	butylpyrazine 5-sec-butyl-2,3-	
31	dimethylpyrazine	1040	dimethylpyrazine	
32		1194		
33	Pentylpyrazine	1192	Pentylpyrazine	
34	2,3-dimetyl-5-pentylpyrazine	1352	2,3-dimetyl-5-pentylpyrazine	
35	Isopentylpyrazine	1157	Isopentylpyrazine	
36	2,3-dimetyl-5-isopentylpyrazine	1317	2,3-dimetyl-5-isopentylpyrazine	
37	(2-methylbutyl)pyrazine	1151	(2-methylbutyl)pyrazine	
38	2,3-dimethyl-5-(2-	1306	2,3-dimethyl-5-(2-	
	methylbutyl)pyrazine		methylbutyl)pyrazine	
39	2-(2-methylbutyl)-2,5,6-	1363	2-(2-methylbutyl)-2,5,6-	
	trimethylpyrazine		trimethylpyrazine	

Figure 2: Table 1 :

2

Name	Definition
MPC03	Molecular path count of order 03
GATS5e Geary	autocorrelation-lag 5/weighted by atomic Sanderson electronegativityies
AEigp	Eigen value distance matrix sum from Polson arizability weight (Barysz matrix)
Qpos	total positive charge
Se	sum of atomic Sanderson electronegativityies
Mp	mean atomic polarizability (scaledon Carbon atom)
X1sol	salvation connectivity index chi-1
DP01	molecular profile no.01
Mor06v	(3D-MORSE-signal 06/weighted by atomic Vander Waals volumes
Tm	T (Total size index/weight atomic masses

Figure 3: Table 2 :

[Ra ()] , Matlab Ra . 2009a.

[Saporta ()] , Gilbert Saporta . 2012. (Régression robuste)

[Pynnönen ()] 'A Report on Least Absolute Deviation Regression with Ordinary Linear Programming'. Seppo Pynnönen , Timosalmi . *Finnish Journal of Business Economics* 1994. 43 (1) p. .

[Dodge and Jureckova ()] *Adaptive Regression*, Y Dodge , J Jureckova . 2000. New York: Springer-Verlag.

[Dodge and Valentin Rousson ()] *Analyses de régression appliquée*, Y Dodge , Valentin Rousson . 2004. (paris)

[Chen et al.] *Analysis of least absolute déviation*, Kani Chen , Zhiliang Ying , Hong Zhang , Lincheng Zhao .

[Damodar et al. ()] *Basic Econometrics Fifth Edition*, N Damodar , Dawn C Gujarati , Porter . 2009.

[Mihara and Enomoto ()] 'Calculation of retention indices of pyrazines on the basis of molecular structure'. S Mihara , N Enomoto . *J. Chromatogr* 1985. 324 p. .

[Stanton and Jurs ()] 'Computer-assisted prediction of gas chromatographic retention indexes of pyrazines'. D T Stanton , P C Jurs . *Anal. Chem* 1989. 61 p. .

[Mihara and Masuda ()] 'Correlation between molecular structures and retention indices of pyrazines'. S Mihara , H Masuda . *J. Chromatogr* 1987. 402 p. .

[Ciuperca ()] *Estimation robuste dans un modèle paramétrique avec rupture*, Gabriela Ciuperca . 2009. Bordeaux.

[Lee et al. ()] 'Evaluation of predictive retention factors for phenolic compounds with QSPR equations'. Seung Ki Lee , Yulia Polyakova , Kyung Row , Ho . *J. Liq. Chromatogr and Rel. Tech* 2004. 27 (4) p. .

[Soumaya] *Indicateurs de la sensibilité de l'estimateur Least Absolute Déviation Assas Paris*, Rekaia Soumaya .

[Faria and Melfi ()] 'Lad regression and nonparametric methods for detecting outliers and leverage points'. S Faria , G Melfi . *Student* 2006. 5 p. .

[Levine ()] I N Levine . *Quantum chemistry. 5*, (New Jersey) 2000. Prentice-Hall. (th ed.)

[Touhami et al. ()] 'Modèle QSRR Hybrides Algorithme Génétique Régression Linéaire Multiple des indices de rétentions de pyrazines en chromatographie gazeuse'. Imen Touhami , Karima Mokrani , Djelloul Messadi . *Lebanese Science Journal* 2012. 13 (1) .

[Machabert ()] *Modèles en très grande dimension avec des outliers. Théorie, simulations, applications" paris*, Tiffany Machabert . 2014.

[Dr and Al -Noor ()] 'Mohammad. Model of Robust Regression with Parametric and Nonparametric Methods'. Nadia H Dr , Asmaa A Al -Noor . *Journal of Mathematical Theory and Modeling* 2013. 3 (5) .

[Parliment and Epstein ()] 'Organoleptic properties of some alkyl-substituted Alkoxy-and alkylthiopyrazines'. T H Parliment , M F Epstein . *J. Agric. Food Chem* 1973. 21 p. .

[Nornadiah Mohd ()] 'Power Comparaisons of shapiro-wilk, Kolmogorovsmornov, lillieffors and Anderson-Darling tests'. Razali Nornadiah Mohd . *Yab Bee Yah*, 2011. 2 p. .

[Kaliszan ()] 'Quantitative relationships between molecular structure and Chromatographic retention'. R Kaliszan . *CRC Crit. Rev. Anal. Chem* 1986. 16 p. .

[Kaliszan ()] *Quantitative structure-chromatographic retention relationships*, R Kaliszan . 1987. New York: J. Wiley.

[Kaliszan ()] *Quantitative structure-chromatographic retention relationships*, R Kaliszan . `http://www.disat.unimib.it22` 1987. New York: J. Wiley. (Moby Digs 1.1)

[Release 16.1, Statistical Software ()] *Release 16.1, Statistical Software*, 2003.

[Saporta ()] *Régression robuste Régression non-paramétrique*, Ndèyeniang-Gilbert Saporta . 2014.

[Dodge ()] *Statistique: Dictionnaire encyclopédique*, Y Dodge . 2004. Springer-Verlag France Paris.

[Masuda et al. ()] 'Synthesis of new pyrazines for flavor use'. H Masuda , Y Misaku , T Shibamoto . *J. Agric. Food Chem* 1981. 29 p. .

[Berlin ()] *The Pyrazine*, G B Berlin . 1982. New York: Wiley-Interscience.

[Buchbauer ()] *Threshold-based structureactivity relationships of pyrazines with bellpepper Flavor*, G Buchbauer . 2000.

[Magnuson et al. ()] 'Topological indices based on neighbor. 26. Symmetry: chemical and biological applications'. V R Magnuson , D K Harriss , S C Basak . *Chemical Applications of Topology and Graph Theory. R.B. King*, (Amsterdam) 1983. Elsevier. p. .

[Masuda and Mihara ()] 'Use of modified molecular connectivity indices to predict retention indices of monosubstituted alkyl, alkoxy, alkylthio, phenoxy and (phenylthio) pyrazines'. H Masuda , S Mihara . *J. Chromatogr* 1986. 366 p. .