Volume : 5 | Issue : 1 | JANUARY 2016 • ISSN No 2277 - 8179

Research Paper

Model of Robust Regression with Least Absolute Deviation Regression and Ordinary Linear Programming Methods of gas chromatographic Retention indices of series compounds of pyrazines

chemistry

KEYWORDS : LAD Regression, Robustness, Outliers, Leverage points,

Fatiha Mebarki	Laboratory of Environmental security and Food; University Badji Mokhtar Annaba,Algeria
Khadidja Amirat	Laboratory of Environmental security and Food; University Badji Mokhtar Annaba,Algeria
Salima Ali Mokhnache	Laboratory of Environmental security and Food; University Badji Mokhtar Annaba,Algeria
Djelloul Messadi	Laboratory of Environmental security and Food; University Badji Mokhtar Annaba,Algeria

ABSTRACT

LAD regression diagnostics offers alternative dicapproaches whose main feature is the robustness. Here a non-parametric method for detecting influential observations is presented and compared with least squares regression method. Was applied to Model, separately, the retention indices of the same set of (89 pyrazines of test and 25 of validation) eluted in turn on OV- 101 and Carbowax-20M, using theoretical molecular descriptors derived from DRAGON Software.

I-Introduction

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 non-linear 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 (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;Tropsha et al..2003).

II.Methodology

II.1.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 5.4[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 Kaliszan.

By employing the software Mobydigs (Todeschini et al., 2009) [21] and by maximizing the coefficient of prédictionQ2 and minimal R2 of S (the error).

II.2.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]:

$$y_i = \beta_0 + \sum_{j=1}^{p-1} \beta_j x_{ij} + \varepsilon_i \qquad (1)$$

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

Volume : 5 | Issue : 1 | JANUARY 2016 • ISSN No 2277 - 8179

squares) [9].

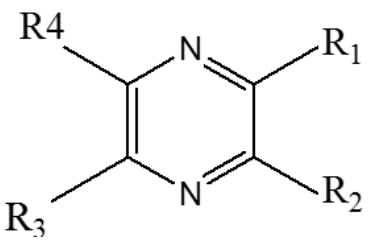
II.2.1.Method of least squares OLS

This one was carried out with the software Minitab 16 [33], method MLR applied to the multiple regression consists in defining the β estimate which minimizes ([9, 17, 18] :

$$\sum ei^2 = \sum (yi - \beta 0 - \sum \beta\, xij)^2 \qquad (2)$$

II.2.2.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 β estimates which minimize [9, 17, the 18]:



(3)

r software Hyperchem
mploying semi-empiri-
Holder 1998) to obtain
lied in this study have

R1: H, alkyl, alkoxy, alkylthio, aryloxy, arylthio, acetyl, chloro.
R2: H, alkyl, chloro, vinyl.
R3: H, alkyl.
R4: H, alkyl.

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.

IV.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 MobyDigs [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.

Table I

Experimentally determined Retention Indices for pyrazines on OV-101 an		
n°	Compounds	ov-101
1	Pyrazine	710
2	Methylpyrazine	801
3	2,3-dimethylpyrazine	897
4	2,5-dimethylpyrazine	889
5	2,6-dimethylpyrazine	889
6	Trimethylpyrazine	981
7	Trimethylpyrazine	1067
8	Ethylpyrazine	894
9	2-ethyl-5-methylpyrazine	980
10	2-ethyl-6-methylpyrazine	977
11	2,5-dimethyl-3-ethylpyrazine	1059
12	2,6-dimethyl-6-ethylpyrazine	1064
13	2,3-dimethyl-5-ethylpyrazine	1066
14	2,3-diethylpyrazine	1065
15	2,3-diethyl-5-methylpyrazine	1137
16	Propylpyrazine	986
17	2-methyl-3-propylpyrazine	1072
18	2,3-dimethyl-5-propylpyrazine	1154
19	2,5-dimethyl-3-propylpyrazine	1142
20	2,6-methyl-3-propylpyrazine	1151
21	Isopropylpyrazine	949
22	2,3-dimethyl-5-isopropylpyrazine	1112
23	Butylpyrazine	1088
24	2-butyl-3-methylpyrazine	1121
n°	Compounds	ov-101
25	3-butyl-3,5-dimethylpyrazine	1184
26	3-butyl-3,6-dimethylpyrazine	1196
27	5-butyl-2,3-dimethylpyrazine	1254
28	Isobutylpyrazine	1043
29	2,3-dimethyl-5-isobutylpyrazine	1200
30	2-isobutyl-3,5,6-trimethylpyrazine	1263
31	sec-butylpyrazine	1040
32	5-sec-butyl-2,3-dimethylpyrazine	1194
33	Pentylpyrazine	1192
34	2,3-dimetyl-5-pentylpyrazine	1352
35	Isopentylpyrazine	1157
36	2,3-dimetyl-5-isopentylpyrazine	1317
37	(2-methylbutyl)pyrazine	1151
38	2,3-dimethyl-5-(2-methylbutyl)pyrazine	1306
39	2-(2-methylbutyl)-2,5,6-trimethylpyrazine	1363

2	Methylpyrazine	801	Methylpyrazine	1235
3	2,3-dimethylpyrazine	897	2,3-dimethylpyrazine	1309
4	2,5-dimethylpyrazine	889	2,5-dimethylpyrazine	1290
5	2,6-dimethylpyrazine	889	2,6-dimethylpyrazine	1300
6	Trimethylpyrazine	981	Trimethylpyrazine	1365
7	Trimethylpyrazine	1067	Trimethylpyrazine	1439
8	Ethylpyrazine	894	Ethylpyrazine	1300
9	2-ethyl-5-methylpyrazine	980	2-ethyl-5-methylpyrazine	1357
10	2-ethyl-6-methylpyrazine	977	2-ethyl-6-methylpyrazine	1353
11	2,5-dimethyl-3-ethylpyrazine	1059	2,5-dimethyl-3-ethylpyrazine	1400
12	2,6-dimethyl-6-ethylpyrazine	1064	2,6-dimethyl-6-ethylpyrazine	1415
13	2,3-dimethyl-5-ethylpyrazine	1066	2,3-dimethyl-5-ethylpyrazine	1421
14	2,3-diethylpyrazine	1065	2,3-diethylpyrazine	1417
15	2,3-diethyl-5-methylpyrazine	1137	2,3-diethyl-5-methylpyrazine	1459
16	Propylpyrazine	986	Propylpyrazine	1374
17	2-methyl-3-propylpyrazine	1072	2-methyl-3-propylpyrazine	1438
18	2,3-dimethyl-5-propylpyrazine	1154	2,3-dimethyl-5-propylpyrazine	1500
19	2,5-dimethyl-3-propylpyrazine	1142	2,5-dimethyl-3-propylpyrazine	1474
20	2,6-methyl-3-propylpyrazine	1151	2,6-methyl-3-propylpyrazine	1493
21	Isopropylpyrazine	949	Isopropylpyrazine	1316
22	2,3-dimethyl-5-isopropylpyrazine	1112	2,3-dimethyl-5-isopropylpyrazine	1431
23	Butylpyrazine	1088	Butylpyrazine	1474
24	2-butyl-3-methylpyrazine	1121	2-butyl-3-methylpyrazine	1459
n°	Compounds	ov-101	Compounds	IR(cw)
25	3-butyl-3,5-dimethylpyrazine	1184	3-butyl-3,5-dimethylpyrazine	1487
26	3-butyl-3,6-dimethylpyrazine	1196	3-butyl-3,6-dimethylpyrazine	1514
27	5-butyl-2,3-dimethylpyrazine	1254	5-butyl-2,3-dimethylpyrazine	1600
28	Isobutylpyrazine	1043	Isobutylpyrazine	1406
29	2,3-dimethyl-5-isobutylpyrazine	1200	2,3-dimethyl-5-isobutylpyrazine	1525
30	2-isobutyl-3,5,6-trimethylpyrazine	1263	2-isobutyl-3,5,6-trimethylpyrazine	1556
31	sec-butylpyrazine	1040	sec-butylpyrazine	1394
32	5-sec-butyl-2,3-dimethylpyrazine	1194	5-sec-butyl-2,3-dimethylpyrazine	1500
33	Pentylpyrazine	1192	Pentylpyrazine	1575
34	2,3-dimetyl-5-pentylpyrazine	1352	2,3-dimetyl-5-pentylpyrazine	1700
35	Isopentylpyrazine	1157	Isopentylpyrazine	1530
36	2,3-dimetyl-5-isopentylpyrazine	1317	2,3-dimetyl-5-isopentylpyrazine	1655
37	(2-methylbutyl)pyrazine	1151	(2-methylbutyl)pyrazine	1527
38	2,3-dimethyl-5-(2-methylbutyl)pyrazine	1306	2,3-dimethyl-5-(2-methylbutyl)pyrazine	1636
39	2-(2-methylbutyl)-2,5,6-trimethylpyrazine	1363	2-(2-methylbutyl)-2,5,6-trimethylpyrazine	1661

40	(2-methyl-3-pentyl)pyrazine	1240	(2-methyl-3-pentyl)pyrazine	1606
41	(2-ethylpropyl)pyrazine	1121	(2-ethylpropyl)pyrazine	1449
42	(1-methylbutyl)pyrazine	1133	(1-methylbutyl)pyrazine	1471
43	2,3-demethyl-5-(2-methylpentyl)pyrazine	1377	2,3-demethyl-5-(2-methylpentyl)pyrazine	1710
44	Hexylpyrazine	1293	Hexylpyrazine	1668
45	Octylpyrazine	1495	Octylpyrazine	1845
46	2-methyl-3-octylpyrazine	1546	2-methyl-3-octylpyrazine	1956
47	2-methyl-5-(2-methylbutyl)-3-octylpyrazine	1923	2-methyl-5-(2-methylbutyl)-3-octylpyrazine	2200
48	2-methyl-6-(2-methylbutyl)-3-octylpyrazine	1962	2-methyl-6-(2-methylbutyl)-3-octylpyrazine	1962
49	Methoxypyrazine	877	Methoxypyrazine	1306
50	2-methoxy-3-methylpyrazine	954	2-methoxy-3-methylpyrazine	1339
51	2-methoxy-5-methylpyrazine	969	2-methoxy-5-methylpyrazine	1358
52	3-ethyl-2-methoxypyrazine	1037	3-ethyl-2-methoxypyrazine	1400
53	3-isopropyl-2-methoxypyrazine	1078	3-isopropyl-2-methoxypyrazine	1400
54	5-isopropyl-3-methyl-2-methoxypyrazine	1170	5-isopropyl-3-methyl-2-methoxypyrazine	1467
55	5-sec-butyl-3-methyl-2-methoxypyrazine	1250	5-sec-butyl-3-methyl-2-methoxypyrazine	1536
56	5-isobutyl-3-methyl-2-methoxypyrazine	1257	5-isobutyl-3-methyl-2-methoxypyrazine	1556
57	3-methyl-2-methoxy-5-(2-methylbutyl) pyrazine	1362	3-methyl-2-methoxy-5-(2-methylbutyl) pyrazine	1664
58	3-methyl-2-methoxy-5-(2-methylpentyl) pyrazine	1444	3-methyl-2-methoxy-5-(2-methylpentyl) pyrazine	1737
n°	Compounds	ov-101	Compounds	IR(cw)
59	Ethoxypyrazine	959	Ethoxypyrazine	1348
60	2-ethoxy-3-methylpyrazine	1029	2-ethoxy-3-methylpyrazine	1385
61	2-ethoxy-5-methylpyrazine	1047	2-ethoxy-5-methylpyrazine	1418
62	2-ethoxy-3-ethylpyrazine	1101	2-ethoxy-3-ethylpyrazine	1439
63	2-ethoxy-3-isopropylpyrazine	1143	2-ethoxy-3-isopropylpyrazine	1431
64	2-ethoxy-5-isopropyl-3-methylpyrazine	1230	2-ethoxy-5-isopropyl-3-methylpyrazine	1500
65	2-ethoxy-5-isobutyl-3-methylpyrazine	1314	2-ethoxy-5-isobutyl-3-methylpyrazine	1584
66	5-sec-butyl-2-ethoxy-3-methylpyrazine	1306	5-sec-butyl-2-ethoxy-3-methylpyrazine	1566
67	2-ethoxy-3-methy-5-(2methylbutyl) pyrazine	1415	2-ethoxy-3-methy-5-(2-methylbutyl) pyrazine	1693
68	(methylthio)pyrazine	1076	2-ethoxy-3-methy-5-(2-methypentyl) pyrazine	1771
69	3-methyl-2-(methylthio)pyrazine	1151	(methylthio)pyrazine	1600
70	5-methyl-2-(methylthio)pyrazine	1163	3-methyl-2-(methylthio)pyrazine	1616
71	3-ethyl-2-(methylthio)pyrazine	1237	3-ethyl-2-(methylthio)pyrazine	1695
72	3-isopropyl-2-(methylthio)pyrazine	1273	3-isopropyl-2-(methylthio)pyrazine	1692
73	3-isopropyl-3-(methylthio)pyrazine	1362	3-isopropyl-3-(methylthio)pyrazine	1737
74	5-sec-butyl-3-methyl-2-(methylthio) pyrazine	1441	5-sec-butyl-3-methyl-2-(methylthio) pyrazine	1800
75	5-isobutyl-3-methyl-2-(methylthio)pyrazine	1446	5-isobutyl-3-methyl-2-(methylthio)pyrazine	1816
76	3-methyl-5-(2-methylbutyl)-2-(methylthio) pyrazine	1552	3-methyl-5-(2-methylbutyl)-2-(methylthio) pyrazine	1941
77	3-methyl-5-(2-methylpentyl)-2-(methylthio) pyrazine	1638	3-methyl-5-(2-methylpentyl)-2-(methylthio) pyrazine	2008
78	(ethylthio)pyrazine	1148	(ethylthio)pyrazine	1635
79	2-ethylthio-3-methylpyrazine	1215	2-ethylthio-3-methylpyrazine	1655
80	2-ethylthio-5isopropyl-3-methylpyrazine	1418	2-ethylthio-5-isopropyl-3-methylpyrazine	1769
81	5-sec-butyl-2-ethylthio-3-methylpyrazine	1494	5-sec-butyl-2-ethylthio-3-methylpyrazine	1832
82	2-ethylthio-5-isobutyl-3-methylpyrazine	1496	2-ethylthio-5-isobutyl-3-methylpyrazine	1843
83	2-ethylthio-3-methyl-5-(2-methylbutyl) pyrazine	1602	2-ethylthio-3-methyl-5-(2-methylbutyl) pyrazine	1951
84	2-ethylthio-3-methylyl-5-(2-methylpentyl) pyrazine	1686	2-ethylthio-3-methylyl-5-(2-methylpentyl) pyrazine	2026
85	Phenoxypyrazine	1415	Phenoxypyrazine	2104
86	2-methyl-3-phenoxypyrazine	1465	2-methyl-3-phenoxypyrazine	2103
87	5-isopropyl-3-methyl-2-phenoxypyrazine	1620	5-isopropyl-3-methyl-2-phenoxypyrazine	2114
88	5-sec-butyl-3-methyl-2-phenoxypyrazine	1694	5-sec-butyl-3-methyl-2-phenoxypyrazine	2173
89	5-isobutyl-3-methyl-2-phenoxypyrazine	1706	5-isobutyl-3-methyl-2-phenoxypyrazine	2209
90	3-methyl-5-(2-methylpentyl)-2-phenoxypyrazine	1807	3-methyl-5-(2-methylpentyl)-2-phenoxypyrazine	2301
n°	Compounds	ov-101	Compounds	IR(cw)
91	(phenylthio)pyrazine	1606	(phenylthio)pyrazine	2400
92	3-methyl-2-(phenylthio)pyrazine	1658	3-methyl-2-(phenylthio)pyrazine	2399
93	5-isopropyl-3-methyl-2-(phenylthio) pyrazine	1806	5-isopropyl-3-methyl-2-(phenylthio) pyrazine	2375
94	5-sec-butyl-3-methyl-2-(phenylthio) pyrazine	1874	5-sec-butyl-3-methyl-2-(phenylthio) pyrazine	2430
95	5-isobutyl-3-methyl-2-(phenylthio)pyrazine	1882	5-isobutyl-3-methyl-2-(phenylthio)pyrazine	2452
96	3-methyl-5-(2-methylbutyl)-2-(phenylthio) pyrazine	1985	3-methyl-5-(2-methylbutyl)-2-(phenylthio) pyrazine	2569
97	3-methyl-5-(2-methylpentyl)-2-(phenylthio) pyrazine	2064	3-methyl-5-(2-methylpentyl)-2-phenylthio) pyrazine	2669
98	Acetylpyrazine	993	Acetylpyrazine	1571
99	2-acetyl-3-methylpyrazine	1061	2-acetyl-3-methylpyrazine	1567
100	2-acetyl-5-methylpyrazine	1093	2-acetyl-5-methylpyrazine	1625

101	2-acetyl-6-methylpyrazine	1089	2-acetyl-6-methylpyrazine	1618
102	2-acetyl-3-ethylpyrazine	1138	2-acetyl-3-ethylpyrazine	1617
103	2-acetyl-3,5-dimethylpyrazine	1153	2-acetyl-3,5-dimethylpyrazine	1629
104	Chloropyrazine	861	Chloropyrazine	1351
105	2,3-dichloropyrazine	1032	2,3-dichloropyrazine	1581
106	2-chloro-3-methylpyrazine	951	2-chloro-3-methylpyrazine	1399
107	2-chloro-3-ethylpyrazine	1044	2-chloro-3-ethylpyrazine	1467
108	2-chloro-3-isobutylpyrazine	1187	2-chloro-3-isobutylpyrazine	1575
109	2-chloro-5-isipropyl-3-methylpyrazine	1173	2-chloro-5-isipropyl-3-methylpyrazine	1505
110	5-sec-butyl-2-chloro-3-methylpyrazine	1256	5-sec-butyl-2-chloro-3-methylpyrazine	1577
111	2-chloro-5-isobutyl-3-methylpyrazine	1264	2-chloro-5-isobutyl-3-methylpyrazine	1600
112	2-chloro-3-methyl-5-(2-methylbutyl) pyrazine	1371	2-chloro-3-methyl-5-(2-methylbutyl) pyrazine	1710
113	2-chloro-3-methyl-5-(2-methylpentyl) pyrazine	1456	2-chloro-3-methyl-5-(2-methylpentyl) pyrazine	1789
114	2-VinylPyrazine	907	2-VinylPyrazine	1392

The definition of each descriptor is given table 2:

Table II
Definitions of Descriptors used in the Retention index Prediction Models [19].

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
Name	Definition
DP01	molecular profile no.01
Mor06v	(3D-MORSE-signal 06/weighted by atomic Vander Waals volumes
Tm	T (Total size index/weight atomic masses

The coefficient of multiple determinations (R2) 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].

IV.1. The best models:
IR(OV-101):(MPC03,X1sol,GATS5e,AEigp,L3e,Qpos);S=20.892, R2=99.30,n=89 compounds.

IR(RWC) : Se, Mp, X1sol,DP01,Mor06v,Tm;S=22.64, R2=99.22,n=89 compounds.

The diagnostic statistics joined together in Table 3 make it possible to make comparisons and to draw several conclusions [21].

Table III
Diagnostiques Statistiques pour les Modèles Sélectionnés

ID	Size	Models	R2	Q2	v	Q2ext	R2adj
OV-101	6	MPC03 X1sol GATS5e AEigp L3e Qpos	99.30	99.12	98.99	96.94	99.24
			SDEP	SDEC	F	s	
			22.448	20.05	1927.2	20.89	

IDx<	Size	Models	R2	Q2	Q2boot	Q2ext	R2adj
CRW-20M	6	Se Mp X1sol DP04 Mor06v Tm	99,2	99	98,92	75,9	99,2
			SDEP	SDEC	F	s	
			24,1	21,7	1740	22,6	

Values of R2 and of R2 (adj) show, each time, quality of adjustment, whereas the very weak differences between R2 and Q2 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 (QBOOT) confirms all at the same time the capacity of internal prediction and the stability of the models.

IV.2. 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].

IV.2.1. Comparison Robust Regression of OLS and LAD:
More particularly we will test 2 methods of estimate for the vector of the Parameters $(\beta_0^*, \beta_1^*, \ldots, \beta_k^*)$:

- 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].

IV.2.1.1. Comparison of hyperplanes of regression:
Column OV-101:
1/LAD:
Y=-48.05-10.14 MPC03+337.87 X1sol -35.78 GATS5e -2.54 AEigp -38.51 L3e -156.88 Qpos (4)

2/OLS :
Y = - 31,2 - 7,77 MPC03 + 300 X1sol - 24,9 GATS5e - 2,31 AEigp - 53,1 L3e - 62,6 Qpos (5)

Column CRW -20M:
1/LAD:
Y= -242, 89 -42, 45 Se+ 687, 45 Mp+ 298,16X1sol+ 205, 42*DP01+ 200,62Mor06v+ 8,04Tm (6)

2/OLS:
Y = - 167 - 42,8 Se + 755 Mp + 320 X1sol + 130 DP01 + 163 Mor06v + 10,7 Tm (7)

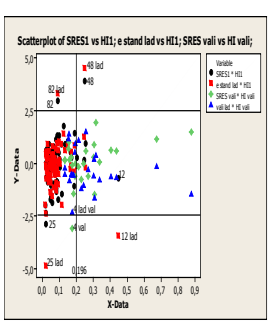
Each equation on each column check the assumptions on the same linear statistical model for Fixes purposes for each method in comparison with the hyperplane calculated by LAD compared to the hyperplane calculated by the method of least squares.

It is noticed that β the calculated OLS are not very different for the regression with β the LAD on the two columns, except, β1the calculated OLS is almost the same ones as for the regression with β1the LAD on column CRW and β4the calculated OLS is almost the same ones as for the regression with β4the LAD on column OV-101.

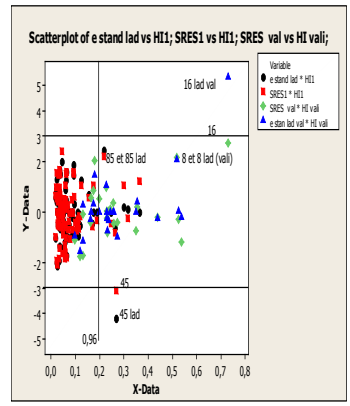
It is thus relevant to remake a checking of the presences of aberrant values by using the following stage (figure 3):

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

IV.2.1.2. Graphical Comparisons of Alternative Regression Models
The field of application was discussed using the diagram of Williams.



Column OV -101



Columns RW -20M

Method LAD and OLS (test, validation)
Fig.1 Diagram of Williams of the residues of prediction standardized according to the lever:
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 25 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)] between the two methods:

Column OV-101:
1/LAD:
y=-48.05-10.14 MPC03+337.87 X1sol -35.78 GATS5e -2.54 AEigp -38.51 L3e -156.88 Qps (8)

2/OLS:
y = - 61,1 - 9,80 MPC03 + 343 X1sol - 35,7 GATS5e - 2,80 AEigp - 40,7 L3e- 160 Qpos (9)

Column CW -20M:
1/LAD:
Y= -242, 89 -42, 45 Se+ 687, 45 Mp+ 298, 16 X1sol + 205,42DP01+ 200,62Mor06v+ 8,04Tm (10)

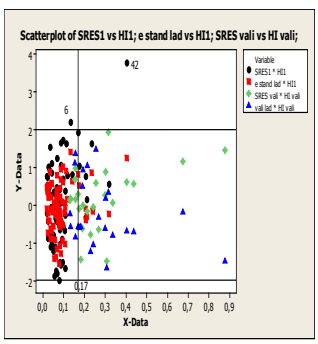
2/OLS:
IR (RCW) = - 192 - 42, 4 Se + 752 Mp + 305 X1sol + 155 DP01 + 156 Mor06v + 13, 0 Tm (11)

It is noticed besides that β the OLS calculate more to approach which for the regression with β the LAD on the two columns into precise (β1 ,β3 and β4) the OLS calculate are almost the

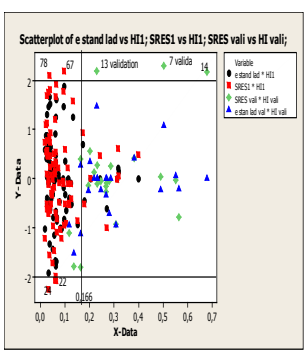
same ones as for the regression with (β1 ,β3 and β4)the LAD and on the same order with (β0 ,β5 and β6) on OV 101 and β1the OLS calculate are almost the same ones as for the regression with β1 the LAD on CRW -20M.

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.2.1.3Graphical Comparisons of Alternative Regression Models



Column OV -101



Columns RW -20M
Method LAD and OLS (test, validation)
Fig.2.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.

V.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:

- The comparison of the equations of the hyperplanes:

L equations of OLS is closer to LAD after elimination of the aberrant points for the β2 (LAD) ≅β2(OLS) and the other coefficient remaining with the same order for column OV-101 Pour the column Crw-20m the β1 (LAD) ≅β1(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, 82,35 24 25) for the column CRW-20M , and to be able to compare them By employing the following stage.

-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.

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

REFERENCE

[1] Stanton, D.T., Jurs, P.C. 1989. Computer-assisted prediction of gas chromatographic retention indexes of pyrazines. Anal. Chem., 61: 1328-1332. [2] Berlin, G .B. 1982 The Pyrazine; Wiley-Interscience: New York. [3] Imen Touhami, Karima Mokrani et Djelloul Messadi. 2012. Modèle QSRR Hybrides Algorithme Génétique Régression Linéaire Multiple des indices de rétentions de pyrazines en chromatographie gazeuse . Lebanese Science Journal, Vol. 13, No. 1. [4] Parliment, T.H., Epstein, M.F. 1973. Organoleptic properties of some alkyl-substituted Alkoxy- and alkylthiopyrazines. J. Agric. Food Chem., 21: 714-716. [5] Kaliszan, R. 1986. Quantitative relationships between molecular structure and Chromatographic retention. CRC Crit. Rev. Anal. Chem., 16: 323-383. [6] Kaliszan, R. 1987. Quantitative structure-chromatographic retention relationships. J. Wiley, New York. [7] Pynnönen, Seppo and Timo Salmi (1994). A Report on Least Absolute Deviation Regression with Ordinary Linear Programming. Finnish Journal of Business Economics 43:1, 33-49. [8] Tiffany Machabert .2014 "Modèles en très grande dimension avec des outliers. Théorie, simulations, applications" paris [9] Dodge, Y. et Valentin Rousson (2004). Analyses de régression appliquée.paris. [10] Kani Chen, Zhiliang Ying, Hong Zhang, and Lincheng Zhao .Analysis of least absolute déviation. [11] Faria, S. and Melfi, G. (2006). Lad regression and nonparametric methods for detecting outliers and leverage points. Student, 5 :265– 272. [12] Gabriela Ciuperca. (2009).Estimation robuste dans un modèle paramétrique avec rupture. Bordeaux. [13] Gilbert Saporta. (2012). Régression robuste. [14] Ndèye Niang- Gilbert Saporta. (2014).Régression robuste Régression nonparamétrique. ¬[15] Dr. Nadia H. AL – Noor and Asmaa A.2013. Mohammad. Model of Robust Regression with Parametric and Nonparametric Methods. Journal of Mathematical Theory and Modeling Vol.3, No.5, [16] Soumaya REKAIA. Indicateurs de la sensibilité de l'estimateur Least Absolute Déviation Assas Paris [17] Dodge, Y. (2004). Statistique : Dictionnaire encyclopédique. Springer-Verlag France Paris. [18] Dodge, Y. and Jureckova, J. (2000). Adaptive Regression. Springer-Verlag New York. [19] Dragon 5.4, http://www.disat.unimib.it [20] Hyperchem 6.03, (Hypercube), http://www.hyper.com. [21] Moby Digs 1.1, http://www.disat.unimib.it [22] Kaliszan, R. 1987. Quantitative structure-chromatographic retention relationships. J. Wiley, New York. [23] Lee, Seung Ki., Polyakova, Yulia. Row, Kyung Ho. 2004. Evaluation of predictive retention factors for phenolic compounds with QSPR equations. J. Liq. Chromatogr and Rel. Tech., 27(4): 629-639. [24] Levine, I.N. 2000. Quantum chemistry. 5 th ed., New Jersey, Prentice-Hall. [25] Magnuson, V.R., Harriss, D.K., Basak, S.C. 1983. Topological indices based on neighbor [26] Symmetry: chemical and biological applications. In: Chemical Applications of Topology and Graph Theory. R.B. King, ed., Elsevier, Amsterdam. 178-191. [27] Masuda, H., Misaku, Y., Shibamoto, T. 1981. Synthesis of new pyrazines for flavor use. J. Agric. Food Chem., 29: 944-947. [28] Masuda, H., Mihara, S. 1986. Use of modified molecular connectivity indices to predict retention indices of monosubstituted alkyl, alkoxy, alkylthio, phenoxy and (phenylthio) pyrazines. J. Chromatogr., 366: 373-377. [29] Mihara, S., Enomoto, N. 1985. Calculation of retention indices of pyrazines on the basis of molecular structure. J. Chromatogr., 324: 428-430. [30] Mihara, S., Masuda, H. 1987. Correlation between molecular structures and retention indices of pyrazines. J. Chromatogr., 402:309-317. [31] Buchbauer, G. 2000. Threshold-based structure-activity relationships of pyrazines with bellpepper Flavor. [32] Matlab Ra 2009a [33] MINITAB, Release 16.1, Statistical Software, 2003 [34] Nornadiah Mohd Razali ,Yab Bee Yah .2011. Power Comparaisons of shapiro-wilk, Kolmogorov-smornov,lillieffors and Anderson-Darling tests,journal of statistique Modelling and analytics .vol 2 No 1:21-33 [35] Damodar N. Gujarati, Dawn C. Porter.2009.Basic Econometrics Fifth Edition