TREATMENT BY ALTERNATIVE METHODS OF REGRESSION GAS CHROMATOGRAPHIC RETENTION INDICES OF 35 PYRAZINES

Fatiha Mebarki, Khadidja Amirat, Salima Ali Mokhnache and Djelloul Messadi

Laboratory of Environmental and Food Safety, University Badji Mokhtar Annaba, Algeria

ABSTRACT

The study treated two closer alternative methods of which the principal characteristic: a non-parametric method (the least absolute deviation (LAD)) and a traditional method of diagnosis OLS.This was applied to model, separately, the indices of retention of the same whole of 35 pyrazines (27 pyrazines with 8 other pyrazines in the same unit) eluted to the columns OV-101 and Carbowax-20M, by using theoretical molecular descriptors calculated using the software DRAGON. The detection of influential observations for non-parametric method (LAD) is a problem which has been extensively studied and offers alternative dicapproaches whose main feature is the robustness .here is presented and compared with the standard least squares regression .The comparison between methods LAD and OLS is based on the equation of the hyperplane, in order to confirm the robustness thus to detect by the meaningless statements and the points of lever and validated results in the state approached by the tests statistics: Test of Anderson-Darling, shapiro-wilk, Agostino, Jarque-Bera, graphic test (histogram of frequency) and the confidence interval thanks to the concept of robustness to check if the distribution of the errors is really approximate.

KEYWORDS

LAD Regression, Robustness, Outliers, Leverage points, tests statistics

1. INTRODUCTION

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 need fondamentaux. Pyrazines are molecule present in a natural way in our environment and which has an interest in multiple fields, in particular in the food like flavor very active. This is why of many studies chemical (Parliment & Epstein, 1973; Masuda *et al.*, 1981; Barlin, 1982; Buchbauer, 2000) on the synthesis and the properties of pyrazines. Their identification is generally done by gas chromatography (CG) by comparing their peaks with those obtained for the standards of the suspectés compounds. The availability of such standards being able to be lacking, the search for other ways of identification is desirable.

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 robustness of method LAD compared to the meaningless statements, and its susceptibility at the point's lever was largely studied in literature (Dodge, 1987;Dodge, 1997).

We propose to method not parametric to detect the influential observations (aberrant and ` action leverage has) with the development of a derive technique of the regression LAD.  The point's lever and the meaningless statements are given by considering suitable disturbances of the whole of source data.  These methods are then compared ` has other know methods.

There is a package of tests of normality Indeed, thanks to the concept of robustness, to check compatibility has the normal law and not to determine the law of distribution, we present the tests statistics of compatibility at the normal law chooses us Ci test if the residues follow a normal law for a risk $\alpha = 5\%$.

The objective of this work aims at using methodology QSRR, in the approach Method LAD/multilinear regression (LAD/OLS), to model the indices of retention of (27) pyrazines reported by Mihara and Enomoto (1985), with 8 other pyrazines taken in the work of Mihara and Masuda (1987) like same unit, 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 and *al*.2003;Tropsha and *al*..2003). We present the tests statistics (Test of Anderson-Darling, shapiro-wilk, Agostino, Jarque-Bera) of compatibility at the normal law for validated the results of the state approached between the two methods for a risk $\alpha = 5\%$.

2.METHODOLOGY

2.1.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 1664 descriptors while functioning on 35 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 digs (Todeschini et al., 2009) [21] and by maximizing the coefficient of prédictionQ^2 and minimal R^2 of S (the error).

2.2.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 squares) [9].

2.2.1. 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.The method LAD applied to the multiple regression consists in defining the β estimates which minimize [9, 17, the 18]:

$$\sum |ei| = \sum |yi\text{-} \beta0\text{-}\sum \beta \; xij| \qquad (2)$$

2.2.2. Method of least squares OLS

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

$$\sum ei^2 = \sum (yi\text{-} \beta0\text{-}\sum \beta \; xij)^2 \qquad (3)$$

3. THE DATA SET

One uses the molecular software Hyperchem 6.03 [20], to represent the molecules, by employing semi-empirical method AM1 (Dewar mud et al., 1985; Holder 1998) to obtain the final geometries. The compounds implied in this study have the general structure 1

R1: H, alkyl, alkoxy, méthylthio, acetyl.
R2: H, alkyl, alkoxy, acetyl, vinyl.
R3:H,alkyl
R4: H, alkyl.
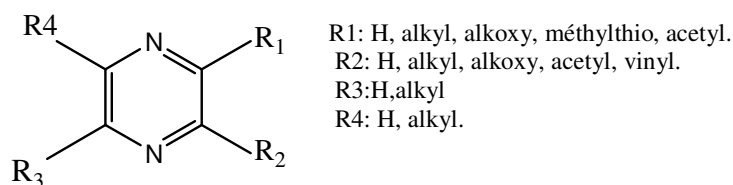
The retention data for the35 compounds chromatographed on the OV-101 and CRW-20M stationary phases were taken from Imen Touhami, Karima Mokrani and Djelloul Messadi [3, 29, and 30] and are listed in table 1.

4. 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 [1, 9].

Table 1:Experimentally determined Retention Indices for pyrazines on OV-101 and Carbowax-20 M

Compound	CW-20M	OV- 101
2-Acetyl 3,6-diMePyrazine	1615.0	1144.0
2-Acetyl 3-EtPyrazine	1617.0	1138.0
2-Methoxy 3-MePyrazine	1339.0	954.0
2-EtPyrazine	1300.0	894.0
2-Et 5-MePyrazine	1357.0	980.0
2-Et 6-MePyrazine	1353.0	977.0
2-Ethoxy 3-MePyrazine	1385.0	1029.0
2-Et 3-MethioxyPyrazine	1695.0	1237.0
2-Et 3-MethoxyPyrazine	1400.0	1037.0
2,3diEt 5-MePyrazine	1459.0	1137.0
2-VinylPyrazine	1392.0	907.0
2-MethoxyPyrazine	1306.0	877.0
2-isoPropyl3-Methoxypyrazine	1400.0	1078.0

2-isoPropyl-3-Methioxypyrazine	1692.0	1273.0
Pyrazine	1179.0	710.0
2-MePyrazine	1235.0	801.0
2,3-diMePyrazine	1309.0	897.0
2,5-diMePyrazine	1290.0	889.0
2,6-diMePyrazine	1300.0	889.0
2,3,5-triMePyrazine	1366.0	981.0
TetraMePyrazine	1439.0	1067.0
2-AcetylPyrazine	1571.0	993.0
2-Acetyl 5-MePyrazine	1625.0	1093.0
2-Acetyl 6-MePyrazine	1618.0	1089.0
2-Acetyl 3,5-diMePyrazine	1629.0	1153.0
2-Acetyl 3-MePyrazine	1567.0	1061.0
2-Ethoxy-3-EtPyrazine	1439.0	1101.0
Acetylpyrazine	1571.0	993.0
2,5-diMe-6-ethylpyrazine	1400.0	1059.0
Butylpyrazine	1474.0	1088.0
5-isopropyl-3Me-2-methoxypyrazine	1467.0	1170.0
Ethoxypyrazine	1348.0	959.0
2-ethoxy-3-ethylpyrazine	1439.0	1101.0
2-Me-3-propylpyrazine	1438.0	1072.0
3-Butyl-2,6-diMepyrazine	1514.0	1196.0

The definition of each descriptor is given on the table (2) [19]:

Table 2. Definitions of Descriptors used in the Retention index Prediction Models [19].

Descriptor	Definition
IC0	Index of the rate of information (symmetry of proximity of order zero)
ATS1p	Autocorrelation of Broto-Moreau of a structure topological of distance 1/balanced by atomic polarizabilities
IVDE	Average rate of information on the equality of the degrees of the tops.
Descriptors	Definition
ESpm04d	Spectral moment 04 of the matrix of adjacency of the edges balanced by the dipole moment
nCconj	A number of combined carbons (sp2).

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].

4.1. The best models

IR (OV-101) :(espmo4, ico, ats1p); S = 13, 1272 R^2 = 99, 0% , n =35 compounds
IR (RCW) :(IVDE, espmo4, ncon); S = 26,1096 R^2= 96,5%, n=35 compounds

Indeed Figure 1 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 σ ^{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 2) is an index of normality. Values of the statistics of Durbin-Watson (Durbin, & Watson, 1951), [d== 2,12225;OV-101/D = 1, 57218;Carbowax-20M] are the greater than higher values given by the tables, respectively for 3 regresses, and any reasonable risk α , which establishes each time the independence of the residues .
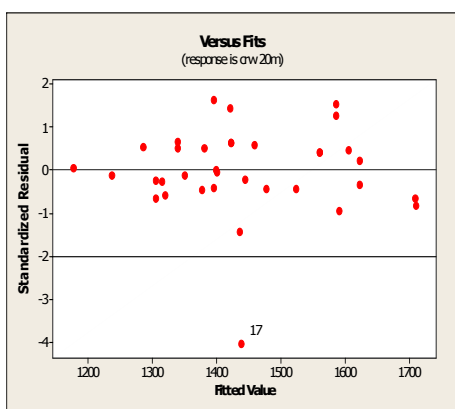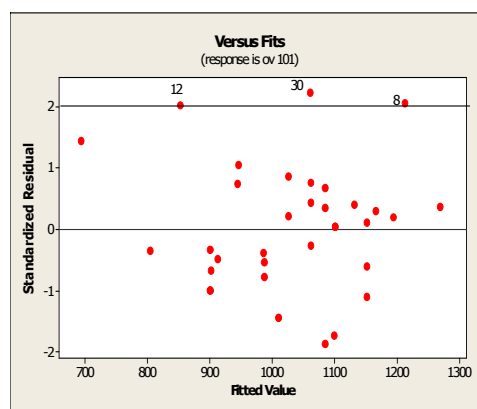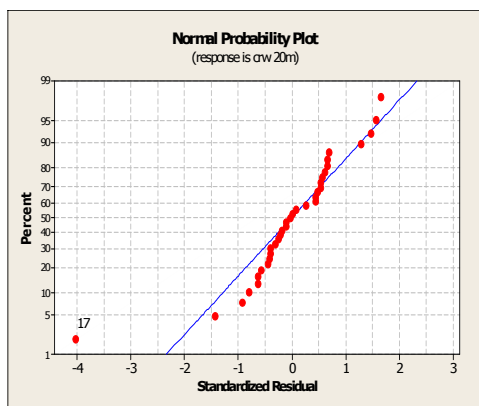

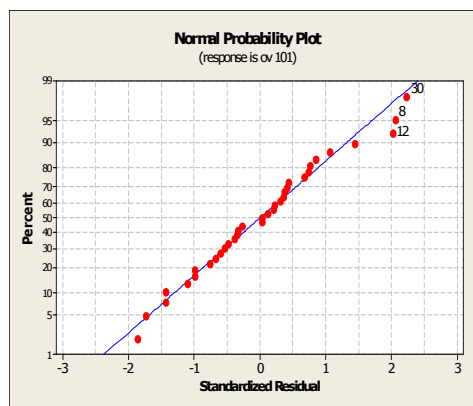The Column RCW -20 M   The Column OV -101

Figure 1 . Plot of the standard residues according to the estimated retention indices

The Column RCW -20 M   The Column OV -101

Figure 2 . Diagram of the normal scores

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

Table **3.** Diagnostic Statistics for the Selected Models

Column	Model	R2	Q2	Q2boot	R2adj	SDEP	SDEC	F	s
OV -101	espmo4 ico ats1p	98.97	98.63	98.42	98.87	14.27	12.35	993.68	13.127
CRW-20M	ivde espmo4 nconj	96.51	95.74	98.00	96.17	27.15	24.57	285.74	26.110

Values of R^2 and R^2_{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 (Q^2_{BOOT}) confirms all at the same time the internal capacity of prediction and the stability of the models.

4.2. 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].**

4.2.1. Comparison Robust Regression of OLS and LAD:

More particularly we will test 2 methods of estimate for the vector of the Parameters $((\beta_0^*, \beta_1^*, ..., \beta_k^*)$:

- 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].

4.2.1.1. Comparison of hyperplanes of regression

The Column CRW -20M :
1/LAD:
y= y= 264.96 -92.587*IVDE+ 253,124*espmo4+ 83.314*nconj   (4)
1/OLS:
y= 242 - 92, 9 IVDE + 259 espmo4 + 66,7 nconj   (5)

The Column OV-101
1/LAD :
y= -1007,5 -9,4* espmo4 + 239,2 ico + 800*ats1p   (6)
2/OLS :
y = - 943 - 5,2 espmo4 + 224 ico + 771 ats1p   (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, $\beta 1$ the calculated OLS is almost the same ones as for the regression with $\beta 1$ the LAD on column CRW-20M.

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.

4.2.1.2. Graphical Comparisons of Alternative Regression Models

Le domaine d'application a été discuté à l'aide du diagramme de Williams.
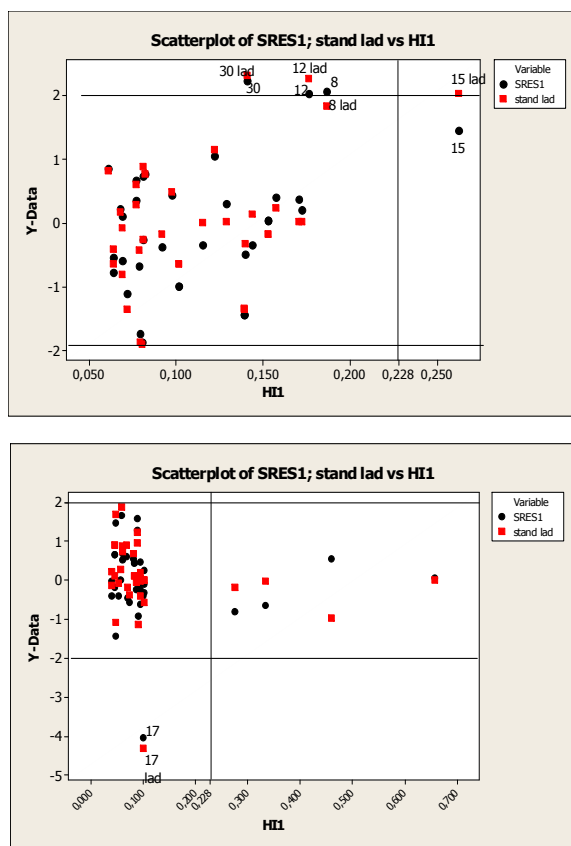
The Column OV -101   The Column CRW-20M

Figure 3. Diagram of Williams of the residues of prediction standardized according to the lever.

As one can see it on the diagrams of Williams, the values hi of all the compounds on column CRW 20 M are lower than the critical value (H * = 0,228) except the item (1, 8, 19, 35) which have $hi > H$ * for the two methods.

On the other hand, for the compounds separated on column OV -101, only one point (15) have *hi* > *H * = 0,228*, others compose, are lower than the critical value. It ya a point (8) *hi > H * = 0,228* on (LAD) inferior has the critical value on (OLS).

In both cases, all the compounds of the residues of prediction standardized and studentuzée lower, in absolute value, with 2 units of standard deviation (σ), except the point (17) for column (CRW) and (30, 15, 12) for column (OV-101) regarded as aberrant points for the two methods;Un only point having a value of Cook > 1 and one value of DFTS lower than the value criticizes $2 * \sqrt{\frac{p}{n}} = 0.67$ for the point (17) which expresses a strong influence, compared to all the compounds on the CRW column. On column OV the 101 all values of Cook are inferiors has 1 as well as the values of DFTS inferiors has the breaking value (= 0.67).

Thus finally the models in which the meaningless statements were removed become :

Column CRW- 20M:

- **LAD:**
- y= 265 -92.587*IVDE+ 253,124*espmo4+   83.314*nconj   (8)
- **OLS:**
- y= 266 - 90,0 ivde + **253** espmo4 + 74,3 nconj   (9)
  Column OV-101 :
- **LAD:**
- y= -1007,5 -9,4 espmo4 + 239,2 ico + 800*ats1p   (10)
  OLS:
- y = - 1008 - 9,3 espmo4 + 232 ico + 804 ats1p   (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 $\beta 0$ *and* $\beta 1$ the OLS calculate are almost the same ones as for the regression with $\beta 0 et \beta 1$ the LAD and on the same order with $\beta 2$ *and* $\beta 3$ on OV 101 and $\beta 0$ *and* $\beta 2$ the OLS calculate are almost the same ones as for the regression with $\beta 0 et \beta 2$ the LAD and on the same order with $\beta 1$ *and* $\beta 3$ on CRW -20M .

The analysis of the residues shows that in this case All the point of LAD and OLS between (-2, 2) the two estimates give good results figure (4):

4.2.1.3. Graphical Comparisons of Alternative Regression Models.

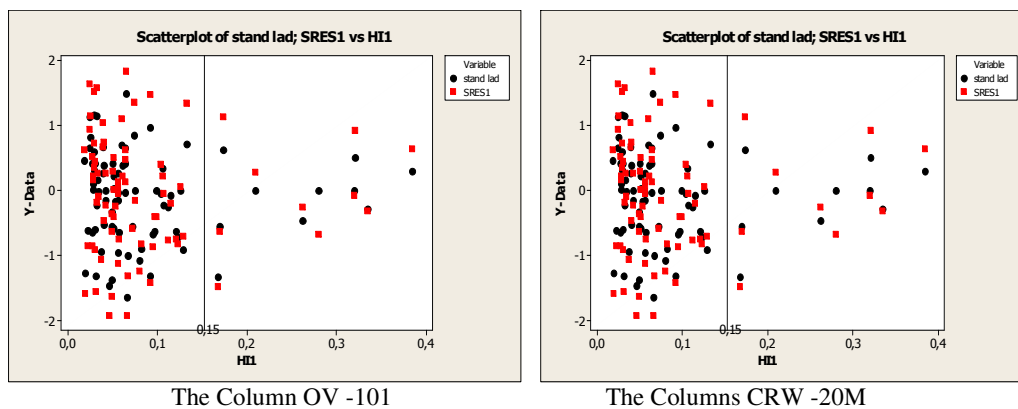The Column OV -101   The Columns CRW -20M

Figure 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 his stability in the presence of the aberrant points.We note however that the least squares estimate is close to the LAD estimate after the elimination of 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 α risk* known as of first species years works us, we will adopt it
risk $α = 5\%$.

4.2.2. 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 ($μ$ the Mean(OV-101:0, CRW-20M:0), $σ$ the variation-type(OV-101:10.35, CRW-20M:14.84) for OLS one applying the same principle with the method LAD but one used (it median (OV-101:-1.57, CRW-20M:0.01) $σ$ variation-type(OV-101:10.26, CRW-20M:15.08)and with the number principal in the state approached to the two columns n=32 .

4.2.2.1.Test statistiques:

4.2.2.1.1. test of Anderson-Darling:

In our work, one finds us that AD [OV- 101:(lad) = 0.351 with value of p >0.250, (OLS) = 0.243 with value of p = 0.747,CRW-20M:(lad) = 0.033 with value of p >0.250, (OLS) = 0.529 with value of p = 0.164] <AD critique=0.752 with p> 0.1 to 5%, the assumption of normality is compatible with our data with the method LAD and OLS.

4.2.2.1.2. / test of Shapiro-Wilk:

For a risk $α = 0.05$, the critical points read in the table of *Shapiro- Wilk* for *N* = 32 is $Wcrit = 0.93$ in our works, *on (OV-101)[W $_{LAD}$ = 0,972 with value of p = 0.165, W_{OLS}= 0.982 with value of p = 0.163] and on CRW -20M[W $_{LAD}$ = 0,99 with value of p = 0.159, W_{OLS}= 0,96 with value of p = 0.168] W > Wcrit*, with p > 0.1with the risk of 5%, the assumption of normality compatible with us is given (normal law).

2-3/ Test of D'Agostino:

For $α = 0.05$, the threshold c0ritic is χ2 0:95(2) = 5.99.In our works, *on (OV-101)[:(W $_{LAD}$ = 5.09 with value of p = 0.078, W_{OLS}= 2,746 with value of p = 0.254, n=32),:] and on CRW -20M[(W $_{LAD}$ = 2.41 with value of p = 0.299,W $_{OLS}$= 1.69 with value of p = 0.429, n=32),] W <Wcrit*, with p > 0.1 with the risk of 5%, the assumption of normality compatible with us is given (normal law).

### 4.2.2.1.3. 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 α = 0.05, the critical point is χ2 0:95(2) = 5.99.In our works, *on (OV-101)[(W $_{LAD}$ = 0.582 with value of p = 0.747, W_{OLS}= 0.008 with value of p = 0.991, n=32),] and on CRW-20M [(W $_{LAD}$ = 1.19 with value of p = 0.551, W_{OLS} = 0,028 with value of p = 0.986, n=32),] W <Wcrit(*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).

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 OLS which explains than give them method LAD is effective and robust para for give method OLS.

4.2.2.2.Tests graphiques :

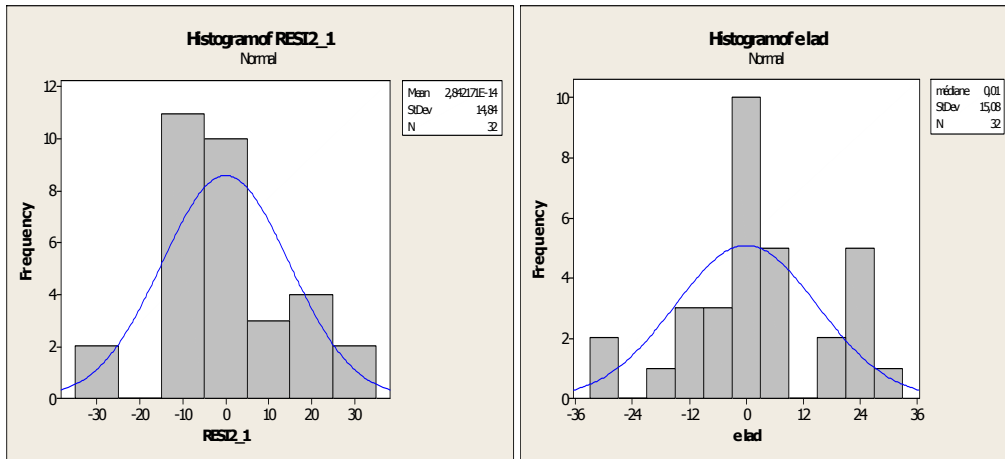### 4.2.2.2.1. Histogramme de fréquence du l'erreur :
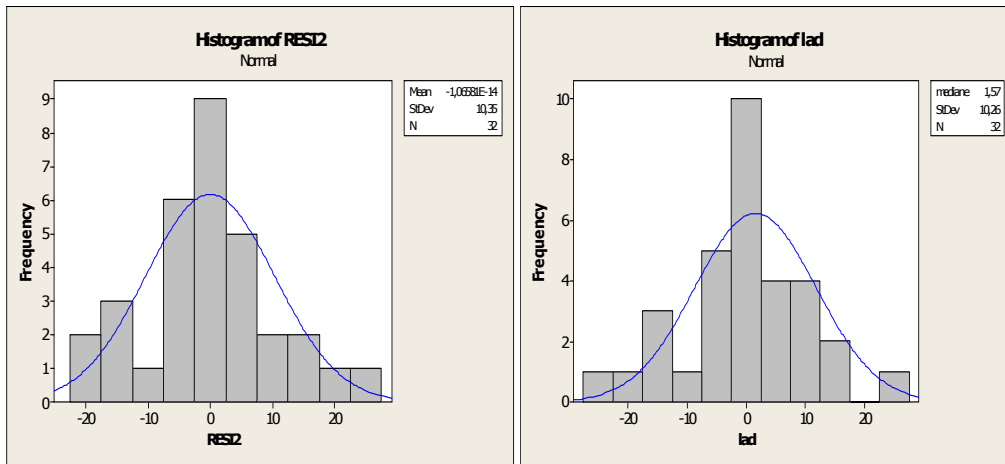
Method LAD   Method OLS

The Column OV-101

Method OLS   Method LAD

The Column CRW-20M

Figure 5. Histogram of frequencies

We observe a standardization of the data, the data are very close to the behavior of the curve of the normal law (all the point between the interval of the curve), and to explain the dependence between the two methods out of both colonnes[33].

4.2.2.3 Interval of confidence:

The confidence interval and the risqe α constitute a complementary approach thus (an approach of estimate) the most used confidence interval is the confidence interval has $100(1 - \alpha) = 95\,\%$.

The Column OV-101:LAD (-3.59, 3.226), OLS (-3.73, 3.73).
The column CRW-20M: LAD (-5.28, 5.30), OLS (-5.35, 5.35).

You can be 95% confident that the 50th percentile for the population is between 3.84295 and 4.58790.

These result is formed L approximate of two method.

5.CONCLUSION

The modeling of the indices of retention of 35 pyrazines eluted out of two columns different OV - 101 and CRW-20M by two methods LAD and OLS is 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 $\beta 1$ (LAD) $\cong \beta 1$(OLS) and the other coefficient remaining with the same order for column OV-101 Pour the column CRW-20M the $\beta 2$ (LAD) $\cong \beta 2$(OLS) and the other coefficient remaining with the same order after the secondary treatments for the checking of the presence of aberrant values (12, 15, 30) on column OV -101 and item 17 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.

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.

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.

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.

REFERENCES

[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 non-paramétrique.

[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

Author

Fatiha Mebarki is currently a PhD Student in Department of Chemistry, University of Badji Mokhtar Annaba, , Algeria. She obtained her D.E.S Degree in Chemistry2002 and her Magister's Degree in analytical chemistry and Environmental (modeling) in 2008 and . Her main research interests include analytical chemistry and Environmental (modeling) from University of Badji Mokhtar Annaba, Algeria.

Amirat Khadjida is PhD student at Badji Mokhtar Annaba, , Algeria her interests are chemo metric bromatology,and analytic chemistry .

Salima Ali Mokhnache is currently a Professor in Department of Chemistry, University of Badji Mokhtar Annaba, , Algeria. She obtained her professor Degree 2009 in Department of Chemistry (the Environmental and Food Safety Laboratory) at University Badji Mokhtar Annaba, Algeria.

Messadi Djelloul is currently a Professor in Department of Chemistry, University of Badji Mokhtar Annaba, , Algeria. He obtained her Baccalaureate (College Saint Augustin) Annaba, , Algeria and her Bachelor of enseignemment Degree in Chemistry-physique from University of Grenoble (France) and her Doctorate of 3 [emme] cycle Degree from University of Saint-Etienne (France), and Doctorate of state in sciences Degree from University of Saint-Etienne (France) in 1981 he became professor in Department of Chemistry at Badji Mokhtar Annaba, , Algeria in 1985.Head of the Environmental and Food Safety Laboratory. Led supported several doctorates. He is co-author of several papers published in international scientific journals. His interests are analytical chemistry, bromatology and chemometric.