Gabriel package: software for education and research

Gilles Olive

Ecole Industrielle et Commerciale de la Ville de Namur, Rue Pépin, 2B, 5000 Namur, Belgium (gilles.olive@eicvn.be)

Résumé

De nos jours, les ordinateurs personnels font partie de notre vie, à la maison, au travail ou bien à l'école. Même si la puissance des ordinateurs augmente, les logiciels sont de plus en plus compliqués, demandant de plus en plus de temps d'apprentissage ou bien pour obtenir un résultat. Il est devenu important d'avoir dans les laboratoires ou à l'école de petits logiciels rapides utilisant des équations empiriques. Nous rapportons ici l'utilisation mais surtout les algorithmes et les équations ainsi que les jeux de paramètres de trois logiciels appartenant à la suite logicielle Gabriel. Le premier, Gabriel Nathalie II, est un calculateur qui a de nombreuses fonctionnalités. Il est possible de calculer tant un pKa que de simuler le déplacement chimique en ^{31}P RMN de molécules du type amino-phosphonate. Par ailleurs, il est également capable de déterminer des valeurs thermodynamiques de fixation de bisphosphonates sur les os humains. Le deuxième, Gabriel Calculatrice Savon, est utilisé pour calculer la quantité d'alcali nécessaire pour faire du savon en fonction de la quantité et de la nature de la matière grasse. Il peut aussi calculer la viscosité de quelques huiles à une température données. Enfin, Gabriel Data Analysis est un logiciel dédié à l'analyse de données. Aujourd'hui, un logiciel pour la recherche peut être utilisé à l'école et un logiciel éducatif peut être utile en recherche.

Mots clés : algorithme, équation, logiciel, analyse de données, savon, pKa, RMN 31P.

Abstract

Today, personal computers are part of our life either at home, at work or at school. Even if the computer power increases, the software are more and more complicated, asking more and more time for learning it or for obtaining a result. It becomes important to have in labs as well as at school small fast software using empirical equations. We report here about the use but more particularly the algorithms, the equations and the data sets of three software belonging to the Gabriel package. The first one, Gabriel Nathalie II, is a calculator with many functionality's. It is possible to calculate the pKa or to simulate the chemical shift in ^{31}P NMR of amino-phosphonate. Moreover it is also able to dertimine the thermodynamics data of bisphosphonate binding to human bones. The second one, Gabriel Calculatrice Savon, is used to calculate the amount of alkali needed for making a soap according to the amount and the nature of fat but also the viscosity of some oils at a desired temperature. And at least Gabriel Data Analysis is a software devoted to data analysis. Today, software for research can be used for educational and educational software can be useful for research.

1. Introduction

It is well established that today, personal computers are part of our life either at home, at work or at school. As the computer power increases, the software becomes harder and harder, requiring more time for learning or for obtaining a result. For example, the use of Gaussian[1], a quantum calculation package, the input creation and more, the calculation requires a long time. Moreover, Gaussian is expensive. Some other software like Tinker[2], Firefly[3] or Facio[4] are free of charge but are more difficult to use even if their authors take the users' wishes into account.

The need of small software using empirical but fast equations becomes urgent in labs as well as at school. Most of this software uses databases patiently collected by the authors.

In this paper we report about the use of three software belonging to the Gabriel package. This software is entirely developed in Visual Basic 6 and can be obtained free of charge upon request. At present some software does not work under Windows Vista or Seven due to some incompatibilities with Visual Basic 6[1]. First, the user has to download and install the software he wants. When first launched, a window containing the license and the history of the software opens. When a new version is out, the user can download the install package or only the executable and then overwrite the old version.

As to the license, this software is not a "freeware" but is "free of charge". They can be used freely but they come with restrictive license agreements. First, the authors are pleased to cite this software (in the About form, it is written how to cite the corresponding one) and to send a copy as pdf file of the publication. Secondly, they are "Scienceware" and are based on the Shareware. Instead of sending money, the users are pleased to send any missing information (like pairing, coefficients for equations, etc ...). For the exact license and to find more information, the users are pleased to click the button View License & History in the About form to read the License & History.txt file.

In this publication, we will also describe the equations, the data sets and the algorithms that are in use in the different software.

[1] An alternative to this problem is to run this software under a Virtual Machine like Virtual PC from Microsoft, Virtual Box, VMware, etc ... In case of difficulties with this alternative, the reader can contact the author for some help. See also the Supporting Information for the list of software running on Vista and Seven and how to install a virtual machine.

As we will see, today, software for research can be used for educational and educational software can be useful for research.

2. Gabriel Nathalie II (GN2)

Pietri *et al.* carried out a very nice work to find the empirical equations for calculating the pKa or the chemical shift in ^{31}P NMR of amino phosphonate.[5-11] Collecting the data (more than 50 points) was very long and needed a great patience. The equations were recalculated several times and improved as new compounds were available. The equations were determined by linear regression from experimental data by using the Program Courbe developed by Blaive.[12] The equation for pKa is (Equation 1)

$$pKa = a_0 + (\sum a_i.n_i) + a_{gem}.n_{gem} \qquad \text{Eq. 1}$$

where n_i represents the number of each substituent of type i, a_0 is a constant value, and $a_{gem}.n_{gem}$ corresponds to the nonlinear steric of a bulky substituent linked in gem position. a_i is a constant increment depending on the substituent. The cyclic alkyl substituent was chosen as a reference and its coefficient a_i was set to 0.

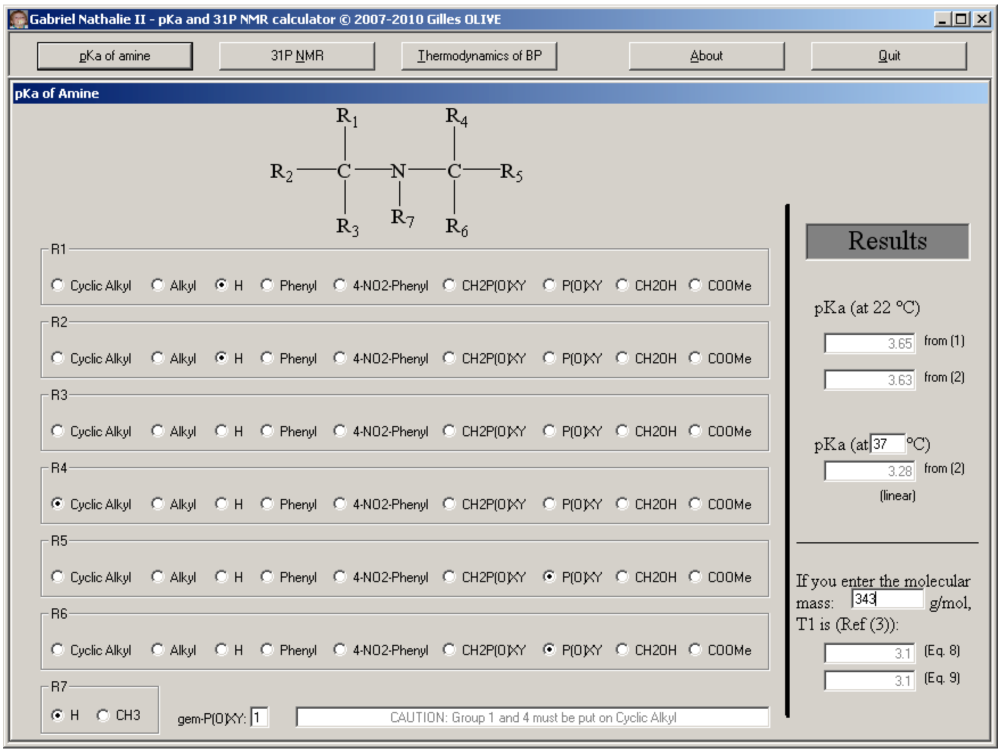
The equations for calculating the chemical shifts (in acidic and in basic medium) or the difference between the two are similar to the equation for calculating the pKa (Equation 2):

$$\delta_{a\text{ or }b} = b_{0a\text{ or }b} + \sum b_{ja\text{ or }b}.n_j \qquad \text{Eq. 2}$$

in which the suffixes a and b mean respectively acid and basic and n_j has the same meaning as above. Alkyl substituents were chosen as reference and their coefficients b_{ja} and b_{jb} were set to 0.

The screen is divided into two unequal parts (Figure 1), the top is used to choose the calculation to be performed. The About button will give the version (3.05 yet) and how to cite this software. Figure 1 shows the screen for pKa calculations. The left part (the biggest one) is used for the introduction of the groups. Seven or nine groups are available according to references [6] and [8]. In the first reference two groups (CH_2OH and COOMe) were not present: choosing one of them will prevent the calculation of the first set of data coefficient.

An important point is that the two sets of data were kept because the second one (on 39 compounds[8]) is not always better than the first one (on 20 compounds[6]). In reference [8], equations for calculating the pKa as a function of the temperature have been introduced but also the possibility to use a tertiary amine (only with a methyl group) instead of a secondary. The attention of the user is drawn to the case of cyclic compound, if R1, R2 or R3 is chosen, then respectively R4, R5 or R6 must be selected as cyclic alkyl. In case of a problem, a message will be sent to the user by the message box (Figure 1). With version above 3.00 of Gabriel Nathalie II, it is possible to calculate the spin lattice relaxation time T_1 in ^{31}P NMR according to two equations[10], both related to the pKa and the molecular mass (Figure 1).



Gabriel Nathalie II - pKa and 31P NMR calculator © 2007-2010 Gilles OLIVE

R1: ○ Cyclic Alkyl ○ Alkyl ● H ○ Phenyl ○ 4-NO2-Phenyl ○ CH2P(O)XY ○ P(O)XY ○ CH2OH ○ COOMe

R2: ○ Cyclic Alkyl ○ Alkyl ● H ○ Phenyl ○ 4-NO2-Phenyl ○ CH2P(O)XY ○ P(O)XY ○ CH2OH ○ COOMe

R3: ○ Cyclic Alkyl ○ Alkyl ○ H ○ Phenyl ○ 4-NO2-Phenyl ○ CH2P(O)XY ○ P(O)XY ○ CH2OH ○ COOMe

R4: ● Cyclic Alkyl ○ Alkyl ○ H ○ Phenyl ○ 4-NO2-Phenyl ○ CH2P(O)XY ○ P(O)XY ○ CH2OH ○ COOMe

R5: ○ Cyclic Alkyl ○ Alkyl ○ H ○ Phenyl ○ 4-NO2-Phenyl ○ CH2P(O)XY ● P(O)XY ○ CH2OH ○ COOMe

R6: ○ Cyclic Alkyl ○ Alkyl ○ H ○ Phenyl ○ 4-NO2-Phenyl ○ CH2P(O)XY ● P(O)XY ○ CH2OH ○ COOMe

R7: ● H ○ CH3

gem-P(O)XY: 1

CAUTION: Group 1 and 4 must be put on Cyclic Alkyl

pKa (at 22 °C)
3.65 from (1)
3.63 from (2)

pKa (at 37 °C)
3.28 from (2)
(linear)

If you enter the molecular mass: 343 g/mol, T1 is (Ref (3)):
3.1 (Eq. 8)
3.1 (Eq. 9)

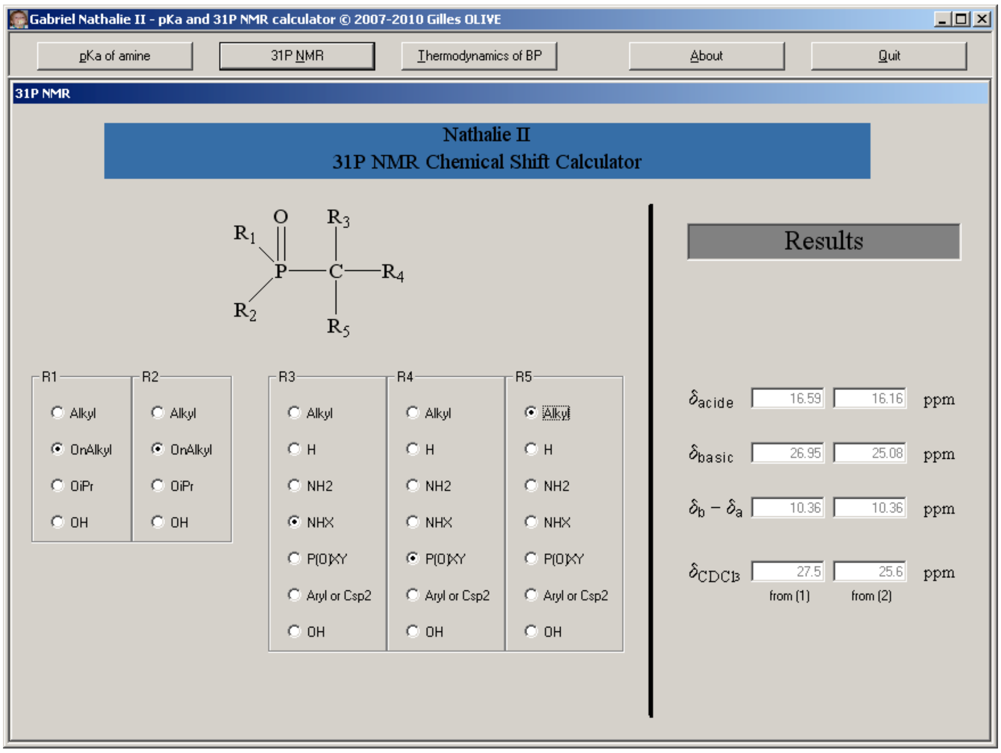
Fig. 1 : Screen of GN2 for calculating pKa's values of amines.[2]

The ^{31}P NMR calculator is based on the same system (Figure 2). The substituent groups are selected with the mouse and the results appear immediately in the right part. Here also, two sets of data, with the reference [6] (on 24 compounds) and [8] (on 44 compounds), are used for each calculation ; none is better than the other when correlated to the experimental values.

[2] The abbreviations have the same meaning as in reference [6]: P(O)XY, phosphonyl substituent.

The difference of acidic and basic chemical shift (δ_b - δ_a) is not the difference between the calculated values but is calculated using another data set[6, 8] of the same type as the acidic or the basic chemical shift.

The value of the chemical shift in the deutareted chloroform was determined statically by the author and was found to be the chemical shift value in the basic medium multiplied by 1.02 if the basic chemical shift is positive or divided by the same value if it is negative (Figure 2).


Gabriel Nathalie II - pKa and 31P NMR calculator © 2007-2010 Gilles OLIVE
pKa of amine
31P NMR
Thermodynamics of BP
About
Quit
31P NMR
Nathalie II
31P NMR Chemical Shift Calculator
O
R1
R2
P
C
R3
R4
R5
Results

R1	R2	R3	R4	R5
○ Alkyl	○ Alkyl	○ Alkyl	○ Alkyl	● Alkyl
● OnAlkyl	● OnAlkyl	○ H	○ H	○ H
○ OiPr	○ OiPr	○ NH2	○ NH2	○ NH2
○ OH	○ OH	● NHX	○ NHX	○ NHX
		○ P(O)XY	● P(O)XY	○ P(O)XY
		○ Aryl or Csp2	○ Aryl or Csp2	○ Aryl or Csp2
		○ OH	○ OH	○ OH

δ_{acide}	16.59	16.16	ppm
δ_{basic}	26.95	25.08	ppm
$\delta_b - \delta_a$	10.36	10.36	ppm
δ_{CDCl3}	27.5	25.6	ppm
	from (1)	from (2)	

Fig. 2 : Screen of GN2 for calculating ^{31}P NMR chemical shift values of phosphorylated compounds.[3]

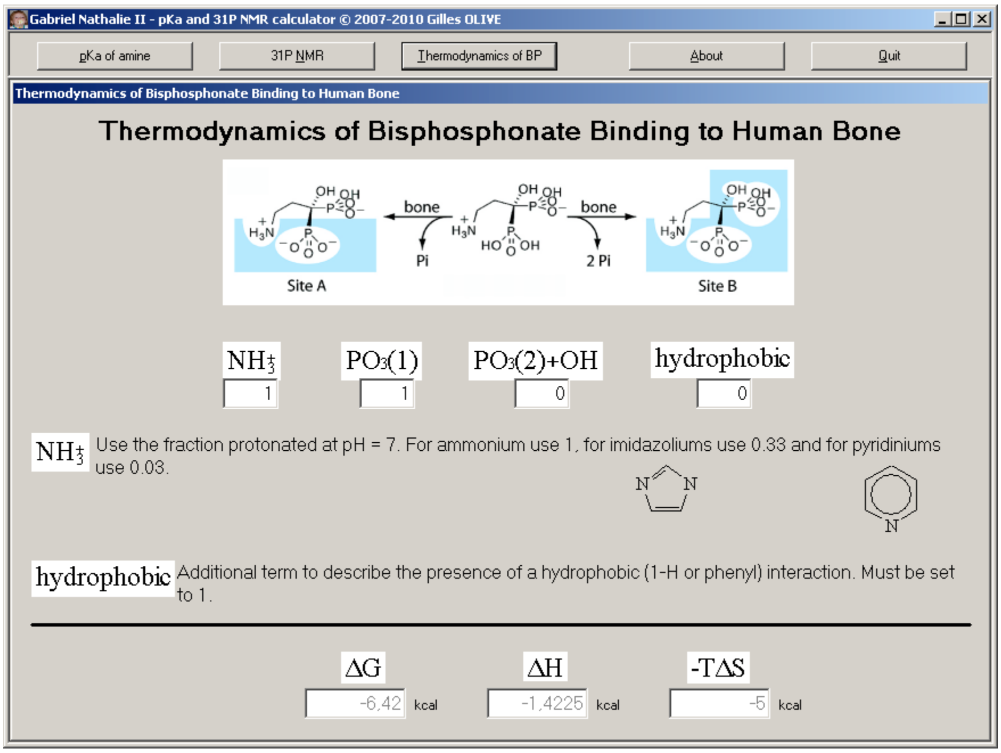
In 2009, the group of Eric Oldfield, in a nice work, found a correlation between the structure and the thermodynamics data of bisphosphonate binding to human bone.[13] The thermodynamics data calculated are the free energy, the enthalpy and the entropy. The equation used are similar to that of calculating the pKa and is a linear combination of variables (on 12 compounds but 18 vectors) (Equation 3):

$$\vec{y} = A\vec{x} \quad \text{where } y_i = \sum a_{i,n}.x_n \qquad \text{Eq. 3}$$

with for example, equation 4

$$\Delta G_{\text{Tot}} = \Delta G_{\text{NH}_3^+} + \Delta G_{\text{PO}_3\,(1)} + \Delta G_{\text{PO}_3\,(2)+\text{OH}} + \Delta G_{\text{hydrophobic}} \qquad \text{Eq. 4}$$

the total free energy of binding is a linear combination of the free energy values for ammonium, phosphonate group number 1, phosphonate group number 2 with an 1-hydroxy group and a general hydrophobic side-chain (hydrogen or phenyl group). As the second phosphonate group that binds to human site B has always an 1-OH group, their contribution was grouped to one only. The equations for enthalpies and for temperature multiplied by entropy are similar to the free energy one. The figure 3 shows how to input the data: just put the number of each group present and the calculations are performed instantly. For the ammonium group, the protonation at pH 7 must be taken and a value of 0.33 and 0.03 respectively for imidazoliums and pyridiniums group has to be mentioned.


Gabriel Nathalie II - pKa and 31P NMR calculator © 2007-2010 Gilles OLIVE
pKa of amine
31P NMR
Thermodynamics of BP
About
Quit
Thermodynamics of Bisphosphonate Binding to Human Bone
bone
bone
Pi
2 Pi
Site A
Site B
N
N
N

Thermodynamics of Bisphosphonate Binding to Human Bone

NH_3^+	$PO_3(1)$	$PO_3(2)+OH$	hydrophobic
1	1	0	0

NH_3^+ Use the fraction protonated at pH = 7. For ammonium use 1, for imidazoliums use 0.33 and for pyridiniums use 0.03.

hydrophobic Additional term to describe the presence of a hydrophobic (1-H or phenyl) interaction. Must be set to 1.

ΔG	ΔH	-TΔS
-6,42 kcal	-1,4225 kcal	-5 kcal

Fig. 3 : Screen of GN2 for calculating thermodynamics values of bisphosphonic acid.

[3] The abbreviations do not have the same meaning as in reference [6]: OnAlkyl, OMe or OEt or OnBu; P(O)XY, phosphonyl substituent; NHX, secondary amino substituent; NH2, primary amino substituent.

It seems that the authors used an another data set because small differences were found between their and our calculations (example on compound 3, site A, -TΔS: -3.9894 kcal by reference [13]; -3.9883 kcal by our software).

3. Gabriel Calculatrice Savon (GCS) (*Gabriel Soap Calculator*)

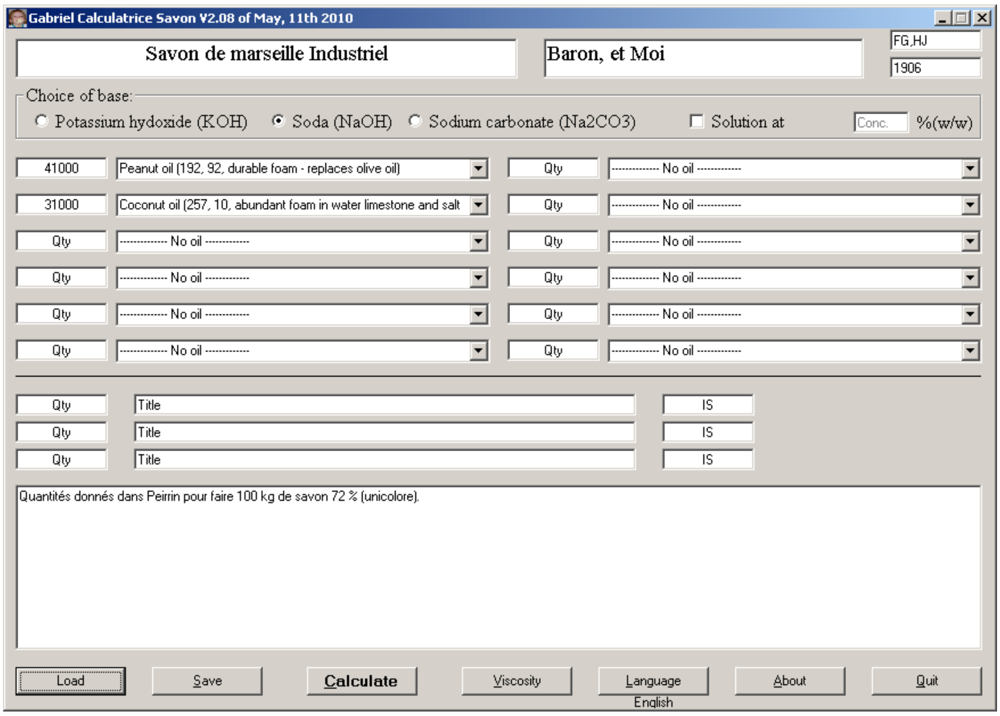
At present, this software is for education only ; at the beginning is was only in French and only six oils are included in the viscosity module. GCS was in French because one of the most cited remarks on Internet forums was that no soap calculator exists in French but most of time in English. The second remark was that the existing soap calculator needed always an Internet access. GCS will soon be usable for research, too. First, we translated it and it is now multi-language as can be seen in figure 4: the proposed languages for all modules are now French, English, Spanish and Dutch (other languages are welcome)[4]. Secondarily we found publications with data set for new oils and new equations for viscosity[14, 15] and an HLB calculator using the Davie's group number will be included[16-20]. Today, this is the 2.08 version of GCS.

Soaps are salts of fatty acids. Making soap is easy. It consists in putting together a fat and an alkali and to boil the mixture for a few hours. Chemically the oils are triglyceride, exactly they are triesters. The action of the alkali is an hydrolysis of the esters to give glycerol (propan-1,2,3-triol) and the salt of fatty acids. It is well known that the quantity of alkali needed for the reaction can be calculated by the Saponification Index (IS in French) or the Saponification Value that is defined as the number of mg of potassium hydroxide needed to neutralize free fatty acids and to hydrolyze the fatty esters in one gram of fat. So this is the total of potassium hydroxide that must be used to make soap with no fatty excess ; it is simple, thanks to the ratio of molecular weight, to know the amount of soda or sodium carbonate. What is very important in such software, is the size of the database of saponification index. Thanks to Internet we have found a database with 89 fats.[21] Of course, verifications with some publications were carried out to avoid any mistakes.

[4] If a file was saved with a Save Version older than 1.2, the language is automatically set to French. Resave the file just after loading and change again the language.

Figure 4 shows the Input Module. At the bottom of the screen, the user can choose the calculation to be performed: load or save a file (in the GCS format), calculate the quantity of base needed, go to the viscosity module, change the language, etc ... The top of the screen is used to introduce the title of the experimentation, the author (for copyright for example), a code (for retrieving a manipulation) and the date.

According to the base (alkali) used, it is possible to carry out different soaps: potassium salts (from potassium hydroxide) give soft soap while sodium salts (from soda or sodium carbonate) give hard soap. In the second line the user can choose which base to use as solid or in solution. In the latter case the solution concentration must be entered in the appropriate textbox (not available if the checkbox is not checked).



Gabriel Calculatrice Savon V2.08 of May, 11th 2010

Savon de marseille Industriel

Baron, et Moi

FG,HJ

1906

Choice of base:

○ Potassium hydoxide (KOH) ● Soda (NaOH) ○ Sodium carbonate (Na2CO3) □ Solution at Conc. %(w/w)

41000	Peanut oil (192, 92, durable foam - replaces olive oil)	Qty	------------ No oil ------------
31000	Coconut oil (257, 10, abundant foam in water limestone and salt	Qty	------------ No oil ------------
Qty	------------ No oil ------------	Qty	------------ No oil ------------
Qty	------------ No oil ------------	Qty	------------ No oil ------------
Qty	------------ No oil ------------	Qty	------------ No oil ------------
Qty	------------ No oil ------------	Qty	------------ No oil ------------

Qty	Title	IS
Qty	Title	IS
Qty	Title	IS

Quantités donnés dans Peirrin pour faire 100 kg de savon 72 % (unicolore).

Fig. 4 : Main screen of GCS.

Next, up to 15 fats can be entered. Three zones are for unknown fat in the database and the name and especially the saponification index must be given. The first twelve areas allow access to the database on fat ; the user enters the weight in grams of the oil and with the combobox he can choose the desired fat among 89 fats. They are divided in five parts: vegetable oil, pure oil compounds, wax, animal fats and vegetable butter. For each entry, the saponification index, Iodine Index or Iodine Value (with these two, the INS (Iodine Number

Saponification Value) can be calculated)[5,6] and some properties of soap made with the oil when they were found. An asterisk following the name of an oil indicates that this oil is present in the Viscosity Module.

And at last, a text box is present to enter some notes and the experimental part.

By pressing the button Calculate, the screen presented in figure 5 will show. At the top, the additional information about the name of the soap, the author, the date, etc ... are recovered from the first form. Follows the quantity of water as an interval calculated by Equation 5

$$\text{quantity of water (ml)} = \text{total amount of fat (g)} * 0.25 + 0.5 \qquad \text{Eq. 5}$$

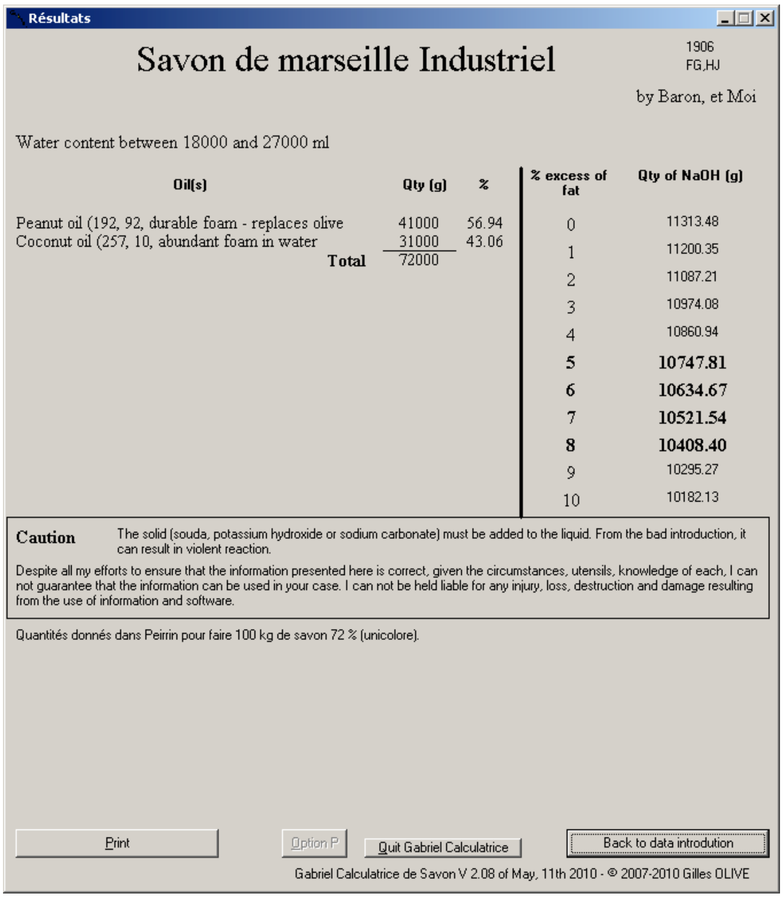
for the lower boundary and by Equation 6 for the upper quantity

$$\text{quantity of water (ml)} = \text{total amount of fat (g)} * 0.375 + 0.5 \qquad \text{Eq. 6}$$

In the right part, the most important one, the quantity of base as a function of the fat excess (up to 10 %) is included while in the left part the user will find a summary of the selected oils. From this point you can print the results, exit GCS or go back to the main form for introducing new data. In fact, the user can print only on the default printer ; soon the printer can be chosen.

[5] If the iodine value was not found, -1 is indicated in the list.

[6] The INS factor is a coefficient which is obtained by removing the iodine value to the saponification index of a fat. It is used to estimate the quality of the soap that will give the mixture of fats and must be around 146.



Savon de marseille Industriel

1906
FG,HJ

by Baron, et Moi

Water content between 18000 and 27000 ml

Oil(s)	Qty (g)	%
Peanut oil (192, 92, durable foam - replaces olive	41000	56.94
Coconut oil (257, 10, abundant foam in water	31000	43.06
Total	72000	

% excess of fat	Qty of NaOH (g)
0	11313.48
1	11200.35
2	11087.21
3	10974.08
4	10860.94
5	**10747.81**
6	**10634.67**
7	**10521.54**
8	**10408.40**
9	10295.27
10	10182.13

Caution The solid (souda, potassium hydroxide or sodium carbonate) must be added to the liquid. From the bad introduction, it can result in violent reaction.

Despite all my efforts to ensure that the information presented here is correct, given the circumstances, utensils, knowledge of each, I can not guarantee that the information can be used in your case. I can not be held liable for any injury, loss, destruction and damage resulting from the use of information and software.

Quantités donnés dans Peirrin pour faire 100 kg de savon 72 % (unicolore).

Gabriel Calculatrice de Savon V 2.08 of May, 11th 2010 - © 2007-2010 Gilles OLIVE

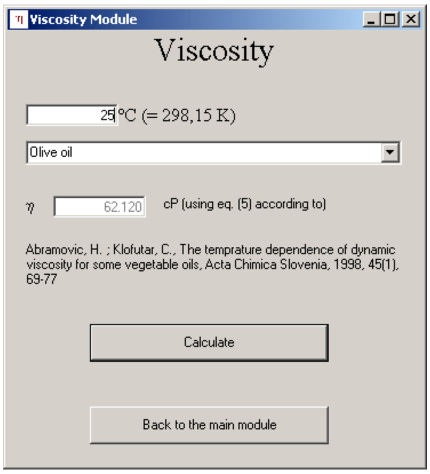
Fig. 5 : Results screen of GCS.

When from the main form the Viscosity button is pressed, the Viscosity Module is activated. In 1998, Abramovic *et al.* published a paper dealing with the temperature dependence of dynamic viscosity for six oils (unrefined sunflower oil, refined sunflower oil, refined corn oil, olive oil, unrefined pumpkin oil, salad oil).[22] In this publication, the authors try to find the best correlation between viscosity and temperature for which they test six different equations, some are modified versions of the Andrade equation. Contrary to what Abramovic[22] wrote, the best fit is not given by their equations 3 or 4 but by equation 5 of the publication. The best equation is given by Equation 7

$$\log \eta = \frac{A}{T} - B \qquad \text{Eq. 7}$$

where T is the absolute temperature in Kelvin, A and B are correlation constants.

The Viscosity Module is easy to use (Figure 6). In the textbox the temperature in degree Celsius is given and in the combobox the oil is chosen between six: unrefined sunflower oil, refined sunflower oil, refined corn oil, olive oil, unrefined pumpkin oil and salad oil. The viscosity value is calculated in centiPoise or mPa.s.



Viscosity Module

Viscosity

25 °C (= 298,15 K)

Olive oil

η 62.120 cP (using eq. (5) according to)

Abramovic, H. ; Klofutar, C., The temprature dependence of dynamic viscosity for some vegetable oils, Acta Chimica Slovenia, 1998, 45(1), 69-77

Fig. 6 : Viscosity screen of GCS.

4. Gabriel Data Analysis (GDA)

Currently GDA is the most important software of this package. The first version, called Gabriel Statistique, was in French, and calculated statistics only on one variable. Due to some bugs and difficult display, in January 2008 the development of Gabriel Statistique was stopped and was replaced by GDA development. In March 2009, the first official version was launched. Today the version 1.17 (Build 1.0) is the latest. To avoid difficult display and to make GDA easier (than Gabriel Statistique), it is a software based on a module's system. We will see each module one after each other, giving equations and algorithms used.

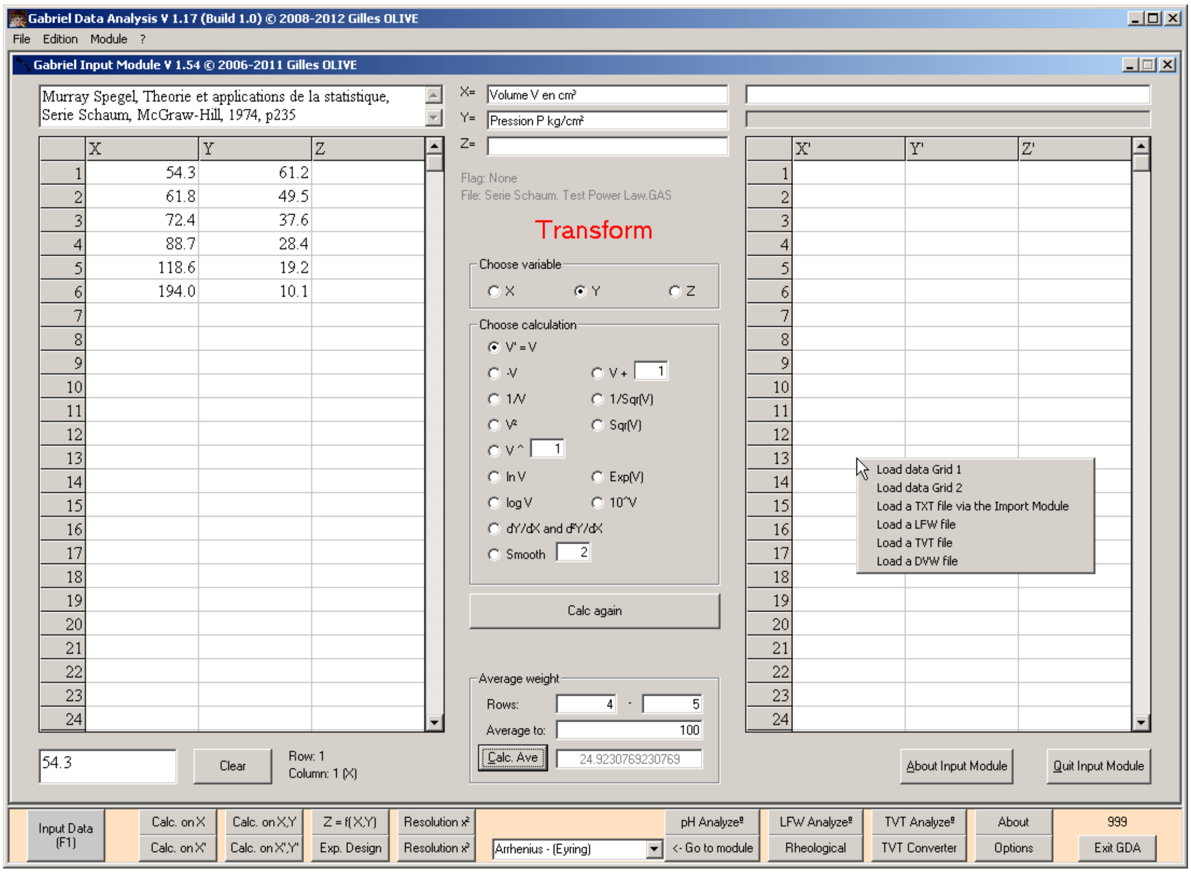
The screen is divided in three parts (Figure 7): the top is the menu, the middle is the module itself and the bottom can be considered as a (tool) bar to choose which module to use. Each module has its own About form and GDA About form also summarizes all information. The desired module is obtained either by pressing the corresponding button, selecting it in the menu (called Module and always present) or using a key combination (see in the Module menu). As the number of modules increases, since the build 4.6 of the version 1.16, some less useful modules are accessible by selecting them in the combobox and pressing the "<- Go to module" button. An ° at the end of the name's module indicates that it needs a special file ; a flag will be up to avoid any problem. Most of time, in a graphical area, a click with the mouse left button will give the coordinates of the plot, according to the units in use.

4.1. Input Module (actual version: 1.54)

This is the first one and the more important module of GDA (Figure 7). With this one, the user can either enter the data, either load data directly from apparatus file. The authors are attentive to the user wishes and are ready to introduce new files format to help users. Several parts are in the screen and the latter revolves around two Grids called respectively Grid 1 (or Major Grid) for the left one and Grid 2 (or Calculation Results Grid) for the right one. From the menu, it is possible to load or save files in several formats: GDA format (that is in fact Gabriel Statistique (.GAS)), ASCII Text file with a complete import module that can import almost all text files[7], STA files which are an old format from the I.E.E.E Module of the homemade Mathematic Solver V 0.16 (on Atari® 520 ST) software, TVT files from Lauda TVT 1 Tensiometer, LFW files from Lauda FW 2 Langmuir-Balance and DVW files from Bohlin CVO120. Some loads can be accessed directly with the right mouse button. CSV (Coma Separate Value) files can be saved and the file can be loaded in Microsoft Excel®. The Edition menu contains normal editing tools like inversion of columns, filling with or without increment a column, deleting or adding a cell, row[8] or column, sorting ascending or descending a column, searching maximum and minimum in each column, imposing a flag and removing some # into the data. The two last deserve more description. To avoid errors in the calculation that could crash the program, when files from apparatus are loaded (or for pH calculations), a flag is set and before launching the corresponding module, a test is carried out to prevent the user's mistake. Grid 2 is used to receive the result of some basic calculations: when a calculation is impossible (the inverse of zero or a logarithm of a negative value for example) to indicate this impossibilities to the user, four # are written in the corresponding cell. Those four # will prevent results of some modules and an option to remove them is implemented in the Edition menu.

[7] With this module you can import Excel files via the CSV format. The actual version of this module is 1.02.

[8] Since version 1.43 of this module, it is possible to delete multiple rows in Grid 2.



Gabriel Data Analysis V 1.17 (Build 1.0) © 2008-2012 Gilles OLIVE

Gabriel Input Module V 1.54 © 2006-2011 Gilles OLIVE

Murray Spegel, Theorie et applications de la statistique, Serie Schaum, McGraw-Hill, 1974, p235

X= Volume V en cm³
Y= Pression P kg/cm²
Z=

	X	Y	Z
1	54.3	61.2	
2	61.8	49.5	
3	72.4	37.6	
4	88.7	28.4	
5	118.6	19.2	
6	194.0	10.1	
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			

54.3

Row: 1
Column: 1 (X)

Flag: None
File: Serie Schaum. Test Power Law.GAS

Transform

Choose variable
○ X ● Y ○ Z

Choose calculation
● V' = V
○ -V ○ V + 1
○ 1/V ○ 1/Sqr(V)
○ V² ○ Sqr(V)
○ V ^ 1
○ ln V ○ Exp(V)
○ log V ○ 10^V
○ dY/dX and d²Y/dX
○ Smooth 2

Average weight
Rows: 4 - 5
Average to: 100
Calc. Ave 24.9230769230769

	X'	Y'	Z'
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			

Arrhenius - (Eyring)

999

Fig. 7 : Input Module screen of GDA.

The middle part contains, above the Grids, a textbox used for comments and three other textboxes to indicate information about the three variables X, Y and Z. The textbox on the right is not accessible to the user and it is used to send information from GDA to the user. The progress bar below has the same meaning. Above the column indication, the file name (when existing) and the eventual flag, are indicated. Between the two grids, the Transform panel ; at the top, the variable that will be changed is selected. Below, the calculations performed: only the last two ones will be detailed, the others do not require special comments. In the X,Y Statistic Module, the same calculations are performed but with different algorithms and they will be explained later. The calculation of the first derivative is depending on the neighborhood of the cell. But before seeing the algorithms as tree, let us present the different equations in which the index 0 represents the line currently being processed, m1 the line just before, m2 the index of the current line less 2, p1 the line just after the current one and p2 the index of the current line plus 2. All calculations are based on finite differences, in particular the forward differences (Equations 8-9), the backward differences (Equations 10-11) and centered differences (Equations 12-13).[23]

$$f'_{+x_0} = \frac{y_{p1} - y_0}{x_{p1} - x_0} \qquad \text{Eq. 8}$$

$$f''_{+x_0} = \frac{y_0 - 2y_{p1} + y_{p2}}{(x_{p2} - x_{p1})(x_{p1} - x_0)} \qquad \text{Eq. 9}$$

$$f'_{-x_0} = \frac{y_0 - y_{m1}}{x_0 - x_{m1}} \qquad \text{Eq. 10}$$

$$f''_{-x_0} = \frac{y_0 - 2y_{m1} + y_{m2}}{(x_0 - x_{m1})(x_{m1} - x_{m2})} \qquad \text{Eq. 11}$$

$$f'_{x_0} = \frac{y_{p1} - y_{m1}}{(x_{p1} - x_0) + (x_0 - x_{m1})} \qquad \text{Eq. 12}$$

$$f''_{x_0} = \frac{y_{m1} - 2y_0 + y_{p1}}{(x_{p1} - x_0)(x_0 - x_{m1})} \qquad \text{Eq. 13}$$

Five types of cells exist: the first, the second, the "middles", the penultimate and the last lines. The algorithms as tree of the five cases are presented in figure 8 to figure 12. In the figures the symbol = means that one of the corresponding cells is empty and the symbol ≠ means that no cell is empty, a # followed by letter(s) means that forth # will be written to the corresponding cell(s). In each case, GDA attempts to calculate the best derivative point.

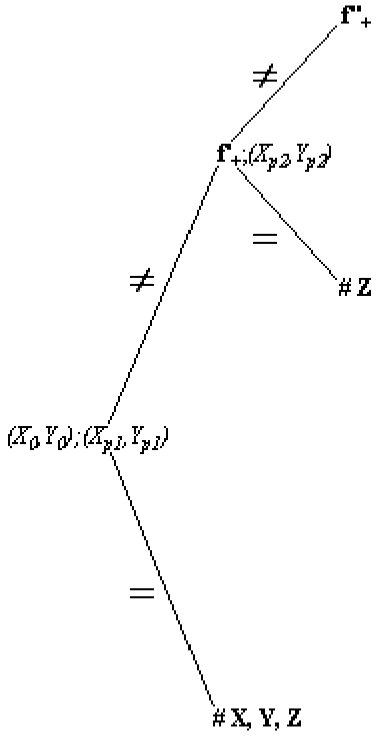


Fig. 8 : Algorithm (as tree) for calculating the best derivative of the First line.

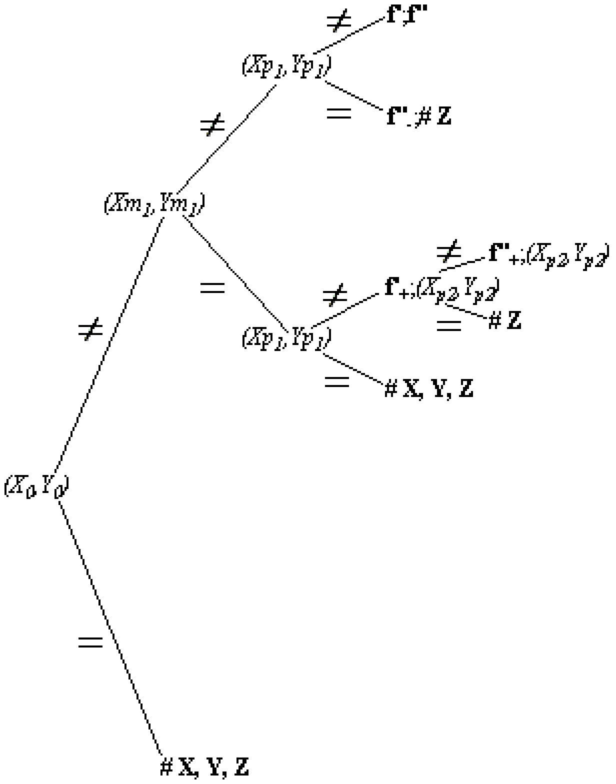


Fig. 9 : Algorithm (as tree) for calculating the best derivative of the Second line.

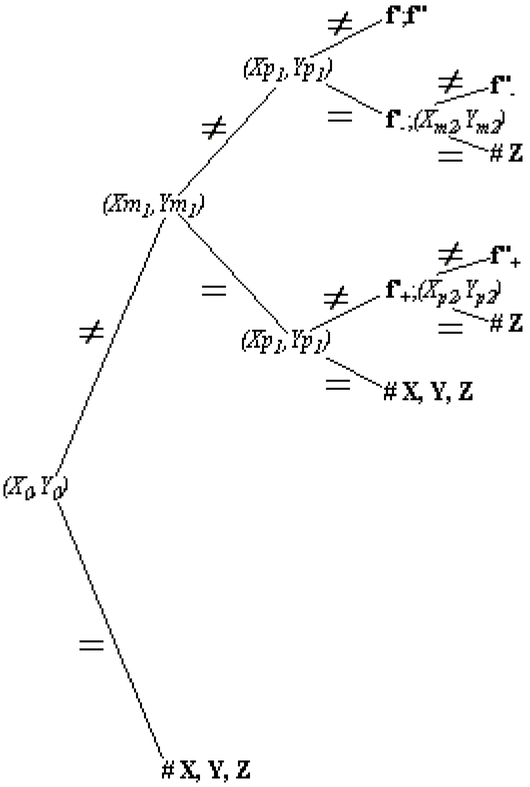


Fig. 10 : Algorithm (as tree) for calculating the best derivative of the "Middles" lines.

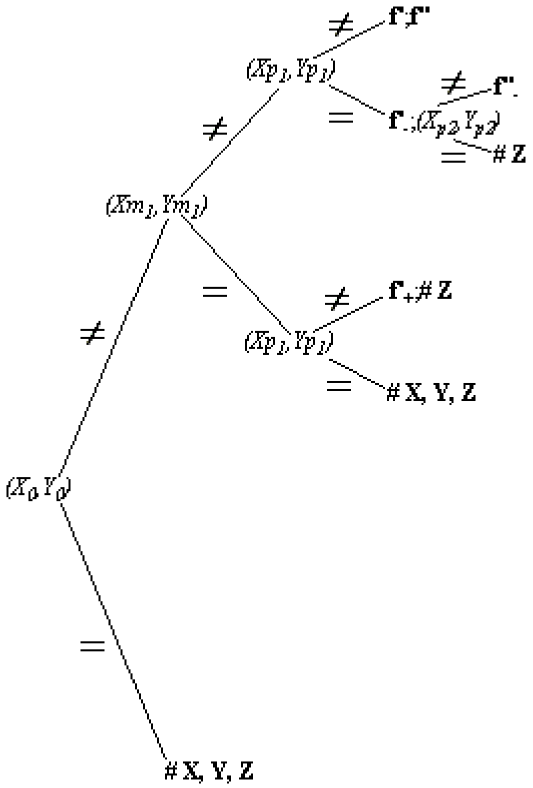


Fig. 11 : Algorithm (as tree) for calculating the best derivative of the "Middles" lines.

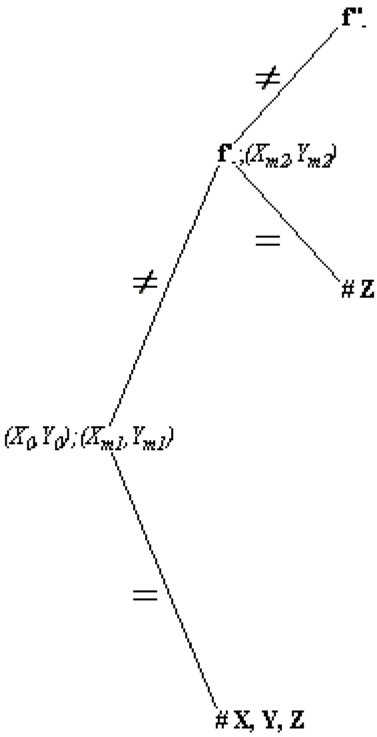


Fig. 12 : Algorithm (as tree) for calculating the best derivative of the Last line.

The second algorithm that will be described here is the algorithm of smoothing. Smoothing the data is used to eliminate noise for example. In the text box, the degree of smoothing is indicated. GDA makes the average of the number of the data before the actual cell corresponding to the degree of smoothing.

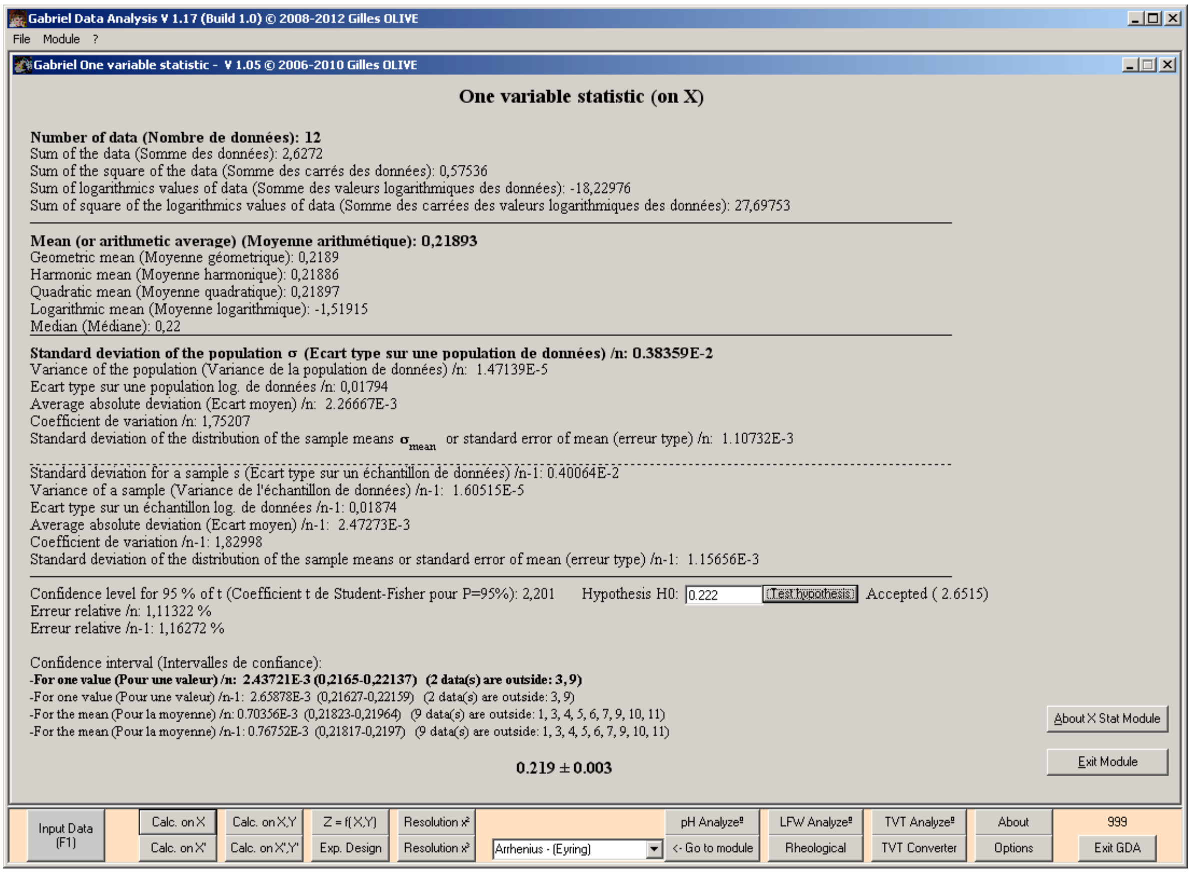
The next part between the two grids is devoted to the average. In many cases, the data is discontinuous and it may be interesting to have the value of Y or Z for a particular value of X. To do this, simply complete the fields within the framework "Average weight", without forgetting to clarify the desired variable within the framework above. The principle is to use a simple rule of three of the difference of the values and add it to the first.

Normally, up to 999 points can be entered ; other capacities of data can be asked to the authors if needed.

When pressing a button in the bottom of the screen, the transfer of the datapoints to the selected module will stop when the first empty cell is found.

4.2. One Variable Statistic Module (actual version: 1.05)

The module can either work on the first column of Grid 1 (X) or Grid 2 (X'). As can be seen in figure 13, this module is an inheritance of Gabriel Statistique: the labels are in English and in French. Only the unusual calculations will be explained here. Figure 13 shows a typical screen results, which is divided in four parts: sums of data, averages, dispersion, Student Fisher analysis.



One variable statistic (on X)

Number of data (Nombre de données): 12
Sum of the data (Somme des données): 2,6272
Sum of the square of the data (Somme des carrés des données): 0,57536
Sum of logarithmics values of data (Somme des valeurs logarithmiques des données): -18,22976
Sum of square of the logarithmics values of data (Somme des carrées des valeurs logarithmiques des données): 27,69753

Mean (or arithmetic average) (Moyenne arithmétique): 0,21893
Geometric mean (Moyenne géometrique): 0,2189
Harmonic mean (Moyenne harmonique): 0,21886
Quadratic mean (Moyenne quadratique): 0,21897
Logarithmic mean (Moyenne logarithmique): -1,51915
Median (Médiane): 0,22

Standard deviation of the population σ (Ecart type sur une population de données) /n: 0.38359E-2
Variance of the population (Variance de la population de données) /n: 1.47139E-5
Ecart type sur une population log. de données /n: 0,01794
Average absolute deviation (Ecart moyen) /n: 2.26667E-3
Coefficient de variation /n: 1,75207
Standard deviation of the distribution of the sample means σ_{mean} or standard error of mean (erreur type) /n: 1.10732E-3

Standard deviation for a sample s (Ecart type sur un échantillon de données) /n-1: 0.40064E-2
Variance of a sample (Variance de l'échantillon de données) /n-1: 1.60515E-5
Ecart type sur un échantillon log. de données /n-1: 0,01874
Average absolute deviation (Ecart moyen) /n-1: 2.47273E-3
Coefficient de variation /n-1: 1,82998
Standard deviation of the distribution of the sample means or standard error of mean (erreur type) /n-1: 1.15656E-3

Confidence level for 95 % of t (Coefficient t de Student-Fisher pour P=95%): 2,201 Hypothesis H0: 0.222 Test hypothesis Accepted (2.6515)
Erreur relative /n: 1,11322 %
Erreur relative /n-1: 1,16272 %

Confidence interval (Intervalles de confiance):
-For one value (Pour une valeur) /n: 2.43721E-3 (0,2165-0,22137) (2 data(s) are outside: 3, 9)
-For one value (Pour une valeur) /n-1: 2.65878E-3 (0,21627-0,22159) (2 data(s) are outside: 3, 9)
-For the mean (Pour la moyenne) /n: 0.70356E-3 (0,21823-0,21964) (9 data(s) are outside: 1, 3, 4, 5, 6, 7, 9, 10, 11)
-For the mean (Pour la moyenne) /n-1: 0.76752E-3 (0,21817-0,2197) (9 data(s) are outside: 1, 3, 4, 5, 6, 7, 9, 10, 11)

0.219 ± 0.003

Fig. 13 : One Variable Statistic Module (results) screen of GDA.

The first part of the screen is easy to understand and will not be discussed here. The second part concerns several averages. The first one, the mean, is well known like the median. The geometric mean G, the harmonic mean H and the quadratic mean Q are respectively defined (and programmed) as (Equations 14-16), in which n is the total number of data[24]

$$G = \sqrt[n]{\prod_{i=1}^{n} x_i} \quad \text{Eq. 14} \qquad H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \quad \text{Eq. 15} \qquad Q = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} \quad \text{Eq. 16}$$

with no negative product of data for G and with no data equal to zero for H as restrictions. The logarithmic mean is defined as (only if all data is positive and non equal to zero) (Equation 17)

$$L = \frac{\sum_{i=1}^{n} \ln x_i}{n} \qquad \text{Eq. 17}$$

The next part of the screen concerns the standard deviations and the variance of the data. This part is divided by two other parts which have the same information, but the upper part concerns a population and n, the number of data, is used in the equations while the lower part concerns a sample and n-1 is used instead.

The equations in use in GDA are for the standard deviation of a population[9] (Equation 18) and for a sample[10] (Equation 19) respectively (in parenthesis how they are defined in reference [24])

$$\mathrm{Sd}_{n} = \sqrt{\frac{\left(n\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}{n^2}} \left(= \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n}} \right) \qquad \text{Eq. 18}$$

and

$$\mathrm{Sd}_{n\text{-}1} = \sqrt{\frac{\left(n\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}{n\,(n-1)}} \left(= \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}} \right) \qquad \text{Eq. 19}$$

and the variance is in the two cases calculated by the Equation 20 ($V = \mathrm{Sd}^2_{n\ \mathrm{or}\ n\text{-}1}$ Eq. 20).

In case of logarithmic data, if all data is superior to zero, the above equations are the same but the sum of data is replaced by the sum of the logarithms of the data. One of the first equations used to have a idea of the dispersion of the data is the average absolute deviation[24] which is (Equation 21)

$$\mathrm{Aad}_{n\ \mathrm{or}\ n\text{-}1} = \frac{\sum_{i=1}^{n}\left|x_i - \bar{x}\right|}{n\ \mathrm{or}\ n-1} \qquad \text{Eq. 21}$$

The variation coefficients are programmed in GDA by the Equation 22

$$\mathrm{VC}_{n\ \mathrm{or}\ n\text{-}1} = \frac{100\,\mathrm{Sd}_{n\ \mathrm{or}\ n\text{-}1}}{\bar{x}} \qquad \text{Eq. 22}$$

[9] which is abbreviated sometime by σ.

[10] which is abbreviated sometime by s.

in which $\overline{x}$ is the arithmetic mean as above and below.

The standard deviation of the distribution of means, also called standard error of mean, is (Equation 23)

$$\mathrm{SE}_{\mathrm{n\ or\ n\text{-}1}} = \frac{\mathrm{Sd}_{\mathrm{n\ or\ n\text{-}1}}}{\sqrt{\mathrm{n}}} \qquad \text{Eq. 23}$$

The last part of the screen is devoted to the Student-Fisher test. The Student test is used for small populations (n < 30). First the corresponding t coefficient for confidence level at 95 % is shown, that can be found at page 344 in reference [24]. The H0 hypothesis will be discussed later. After, the relative error follows (Equation 24)

$$\mathrm{RE}_{\mathrm{n\ or\ n\text{-}1}} = \frac{100 \times \mathrm{Sd}_{\mathrm{n\ or\ n\text{-}1}} \times \mathrm{t}}{\overline{\mathrm{x}}} \qquad \text{Eq. 24}$$

Four confidence levels are then given, respectively for one value and for the mean for n and n-1 values in the two cases. GDA gives respectively for each line the value that allows the calculations of the limits of the confidence, the lower and the upper boundaries of the limits of the confidence and the number of data and which data is outside those limits. The limits of the confidence is described by Equation 25

$$\overline{\mathrm{x}} \pm \frac{\mathrm{Sd}_{\mathrm{n\ or\ n\text{-}1}} \times \mathrm{t}}{\sqrt{\mathrm{n}} \text{ or } \sqrt{\mathrm{n - 1}}} \qquad \text{Eq. 25}$$

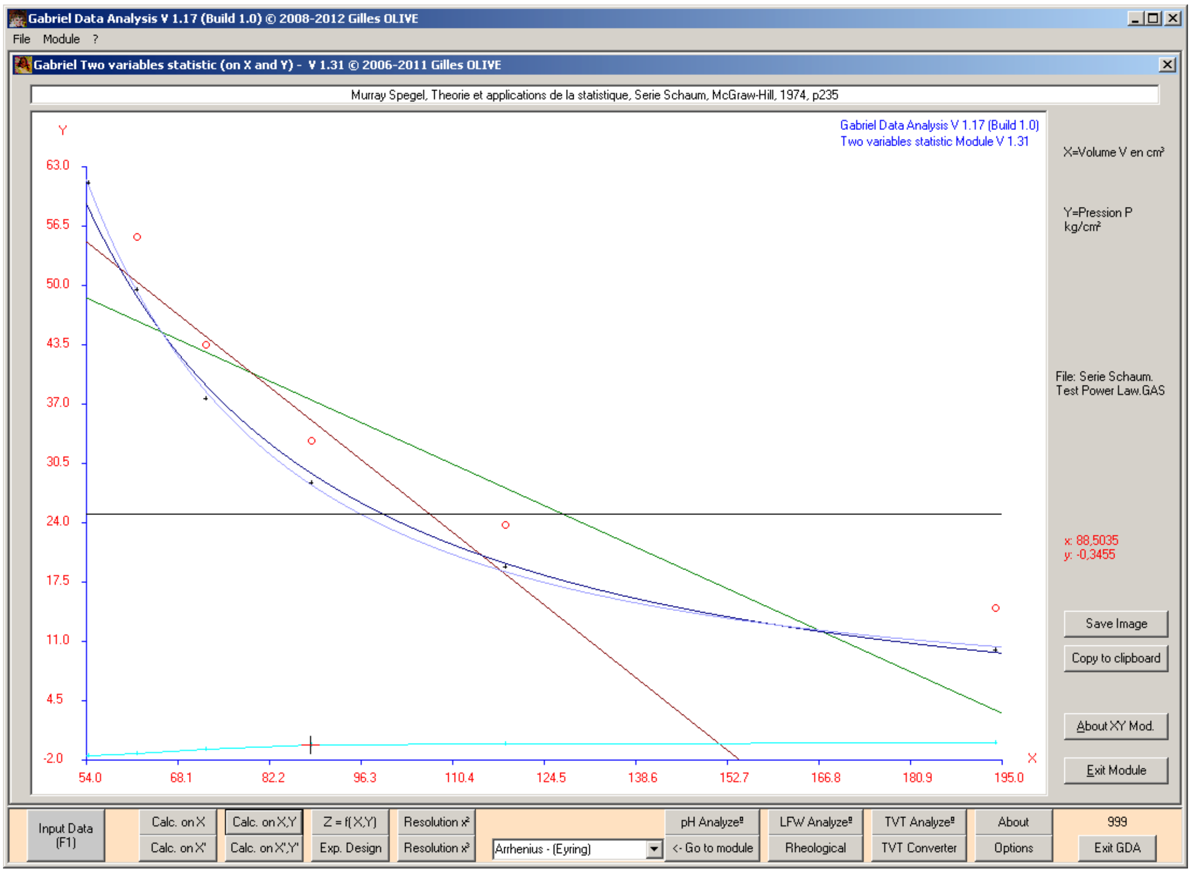
The last part of the equation above is the value given by GDA for calculating the two limits. The results are then given by the mean followed by the "Student error", GDA will give the study of the data to this format: the good number of digit according to the majored value of the error.

A null hypothesis H0 can be tested. First the user enters in the text box the value he wants to test and after pressing on the "Test hypothesis" button the result, accepted or rejected, is given, followed in parenthesis, the value of the Hypothesis Calculated t. As a reminder, the hypothesis is accepted if the Hypothesis Calculated t is greater than t.[24] The Hypothesis Calculated t is calculated according to reference [24] and Equation 26

$$t_{HC} = \frac{\left|\overline{x} - \text{Value to be tested}\right|}{Sd_n}\sqrt{n-1} \qquad \text{Eq. 26}$$

4.3. Two Variables Statistic Module (actual version: 1.31)

This module is the biggest part of GDA and one of the reasons[11] of GDA, the use of Microsoft Excel® is not easy for finding the best correlation of data. This module, like the last described module, can either work from the first (X,Y) or from the second (X',Y') grid. When this module is called, two screens appear. The first is the graphical window (Figure 14) and the second is the equation panel (Figure 15).



Gabriel Data Analysis V 1.17 (Build 1.0) © 2008-2012 Gilles OLIVE

Gabriel Two variables statistic (on X and Y) - V 1.31 © 2006-2011 Gilles OLIVE

Murray Spegel, Theorie et applications de la statistique, Serie Schaum, McGraw-Hill, 1974, p235

Gabriel Data Analysis V 1.17 (Build 1.0)
Two variables statistic Module V 1.31

X=Volume V en cm³

Y=Pression P kg/cm²

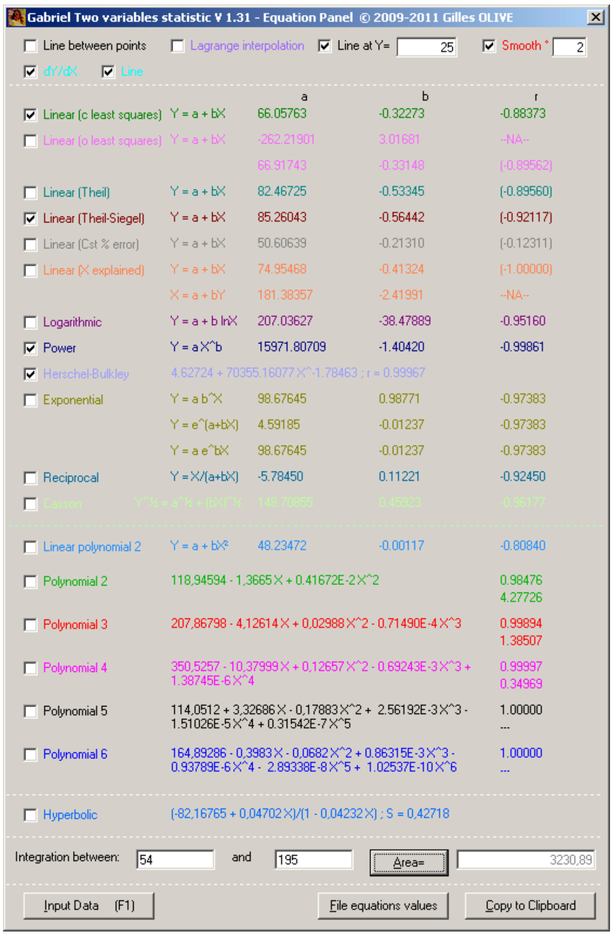
File: Serie Schaum. Test Power Law.GAS

x: 88,5035
y: -0,3455

Arrhenius - (Eyring)

Fig. 14 : Graphical window of the Two Variables Statistic Module of GDA.

A second window is needed, due to the number of equations (more than 20), because we want to have the biggest graphical area to see and be able to choose the best correlation. Figure 14 shows a typical screen: on top, the title (which is the text in the comment text box of the Input Module), in the right part the information (from the Input Module) on the variables X and Y, three buttons to save or copy the graphic and the about of this module, and in the center part the plot itself. In this plot, as a cross, the data are drawn as well as the curves that are chosen in the equation panel (Figure 15). In Figure 15 it is easy to choose what the user wants to draw by selecting the corresponding checkboxes and we will discuss now all calculations.



Gabriel Two variables statistic V 1.31 - Equation Panel © 2009-2011 Gilles OLIVE

□ Line between points □ Lagrange interpolation ☑ Line at Y= 25 ☑ Smooth * 2

☑ dY/dX ☑ Line

		a	b	r
☑ Linear (c least squares)	Y = a + bX	66.05763	-0.32273	-0.88373
□ Linear (o least squares)	Y = a + bX	-262.21901	3.01681	--NA--
		66.91743	-0.33148	(-0.89562)
□ Linear (Theil)	Y = a + bX	82.46725	-0.53345	(-0.89560)
☑ Linear (Theil-Siegel)	Y = a + bX	85.26043	-0.56442	(-0.92117)
□ Linear (Cst % error)	Y = a + bX	50.60639	-0.21310	(-0.12311)
□ Linear (X explained)	Y = a + bX	74.95468	-0.41324	(-1.00000)
	X = a + bY	181.38357	-2.41991	--NA--
□ Logarithmic	Y = a + b lnX	207.03627	-38.47889	-0.95160
☑ Power	Y = a X^b	15971.80709	-1.40420	-0.99861
☑ Herschel-Bulkley	4.62724 + 70355.16077 X^-1.78463 ; r = 0.99967			
□ Exponential	Y = a b^X	98.67645	0.98771	-0.97383
	Y = e^(a+bX)	4.59185	-0.01237	-0.97383
	Y = a e^bX	98.67645	-0.01237	-0.97383
□ Reciprocal	Y = X/(a+bX)	-5.78450	0.11221	-0.92450
□ Casson	Y^½ = a^½ + (bX)^½	148.70855	0.45923	-0.96177
□ Linear polynomial 2	Y = a + bX²	48.23472	-0.00117	-0.80840
□ Polynomial 2	118,94594 - 1,3665 X + 0.41672E-2 X^2			0.98476 4.27726
□ Polynomial 3	207,86798 - 4,12614 X + 0,02988 X^2 - 0.71490E-4 X^3			0.99894 1.38507
□ Polynomial 4	350,5257 - 10,37999 X + 0,12657 X^2 - 0.69243E-3 X^3 + 1.38745E-6 X^4			0.99997 0.34969
□ Polynomial 5	114,0512 + 3,32686 X - 0,17883 X^2 + 2.56192E-3 X^3 - 1.51026E-5 X^4 + 0.31542E-7 X^5			1.00000 ---
□ Polynomial 6	164,89286 - 0,3983 X - 0,0682 X^2 + 0.86315E-3 X^3 - 0.93789E-6 X^4 - 2.89338E-8 X^5 + 1.02537E-10 X^6			1.00000 ---
□ Hyperbolic	(-82,16765 + 0,04702 X)/(1 - 0,04232 X) ; S = 0,42718			

Integration between: 54 and 195 Area= 3230,89

Fig. 15 : Equations Panel of the Two Variables Statistic Module of GDA.

"Line between points" and "Line at Y =" will not be discussed here as they are easy to understand, like the Lagrange interpolation that is well known and often described[23] and is schematically calculated, first searching the Lagrange's polynomial $L_j(x)$ (Equation 27) and,

[11] The second reason is that there is no easy software for kinectics and rheological treatment as they use more than two non linear variables and need sophisticated method like the Levenberg-Marquardt algorithm.

secondly the polynomial $P_n(x)$, which is the interpolation's polynomial, calculated using Equation 28

$$L_j(x) = \prod_{\substack{k=1 \\ k \neq j}}^{n} \frac{x - x_k}{x_j - x_k} \quad \text{Eq. 27} \qquad\qquad P_n(x) = \sum_{j=1}^{n} y_j\, L_j(x) \quad \text{Eq. 28}$$

in which x is the value for x to be interpoled.

Smoothing the data according to the value of the degree of smoothing in the textbox uses the same algorithm as the one described in the Input Module for the y's values only. Until version 1.23 of this module, the smoothed data were accessible by pressing the button "Smooth": all equations were erased and the graphic plot showed only the smoothed data. Now the button is replaced by a checkbox and the smoothed data are added as red circles to the plot of the data and the desired functions. Because there are no recalculations of equations, they are given in the panel for raw data ; in case the user wants the equations for smoothed data, he must use the Input Module for such calculations.

Pressing the button "Area=" at the bottom of the panel, the area will be calculated, using the trapezes method (Equation 29) between the two indicated boundaries a and b (textboxes) with n intervals.[23] In fact GDA makes the sum of the absolute values and the result appears in the textbox on the right side.

$$\int_a^b f(x)\,dx = \sum_{i=1}^{n} \frac{(f_{i-1} + f_i)}{2}\,\Delta x_i \qquad \text{with } \Delta x_i = x_i - x_{i-1} \qquad \text{Eq. 29}$$

Below, there are three buttons. The first on the left side is to go back to the Input Module, "File equations values" button will create a CSV file containing for all values of x (on the screen) all y values according to the selected functions in this panel (see Figure 15), while "Copy to Clipboard" will copy all equations and correlation coefficients to clipboard which can be pasted to any word processing software.

A checkbox is present before the equations and when one is selected, the corresponding function is plotted to the graphical window as shown in Figure 14. Now all algorithms used to obtain the equations will be discussed here.

The first one is the first derivative. As already explained in the Input Module section and as the data are continuous in this module, the algorithm used here is a more simple algorithm than the one used in the Input Module. The first point is always calculated using the forward differences (Equation 8), the last point using the backward differences (Equation 10) and between by the centered differences (Equation 12) with a small difference at the denominator where x_{p1} - x_{m1} is directly calculated instead of (x_{p1} - x_0) + (x_0 - x_{m1}). The scale in the graphical window is recalculated and lines between the first derivative point can be plotted in light blue.

The classical least square (CLS) method for finding the best linear (y = a + b x) fit is well documented[24-26] and is programmed in GDA as follows (Equations 30-32) including r, the correlation coefficient, that gives one of the information of the quality of the fit.

$$a = \frac{\sum_{i=1}^{n} y_i - b \sum_{i=1}^{n} x_i}{n} \qquad \text{Eq. 30} \qquad \text{with} \qquad b = \frac{n \sum_{i=1}^{n} x_i\, y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad \text{Eq. 31}$$

$$r = \frac{n \sum_{i=1}^{n} x_i\, y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\left(n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right)\left(n \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right)}} \qquad \text{Eq. 32}$$

Unfortunately, the classical least square regression suffers several drawbacks that will be discussed now.[26-28]

One of these drawbacks is that the classical least square technique "minimizes only the squared distance parallel to the y axis between the experimental points and the fitting line".[26] Moreover the x's values are not always free of errors or uncertainty.[26] Furthermore, in the CLS regression there is no symmetric treatment of the data x and y series preventing the transformation of a fit y over x in x over y. The best method to avoid these two drawbacks is the use of the orthogonal least square (OLS) technique which minimizes the sum of quadratic distances to the equation line.[26] This has been introduced in GDA according to equations 33 to 35 using a temporary variable as described in reference [26].

$$B=\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2-\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}{2\sum_{i=1}^{n}\left(\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)\right)} \qquad \text{Eq. 33}$$

$$a_{\pm}=\bar{y}-b_{\pm}\,\bar{x} \quad \text{Eq. 34} \qquad \text{with} \qquad b_{\pm}=-B\pm\sqrt{B^2+1} \quad \text{Eq. 35}$$

As the reader can see, the OLS regression gives two equations: the one wanted and the orthogonal one to it. The two fits are plotted if requested.

One of the actual problems is to find an efficient coefficient r that gives the quality of the fit. After finding some in page 244 of reference [24], the Equation 36, calculating correlation coefficient r, has been chosen for OLS, Thiel, Theil-Siegel, assuming a constant percent error and X as explained variable linear regressions. In extreme cases, results are strange (r can be greater than one). TAKE CARE WITH THESE RESULTS. For those of which the correlation coefficient is not efficient, r will be written between parenthesis in the equation panel.

$$r=\mathrm{sign}(b)\times\sqrt{\left|1-\frac{\sum_{i=1}^{n}y_i^2-a\sum_{i=1}^{n}y_i-b\sum_{i=1}^{n}x_i\,y_i}{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}\right|} \qquad \text{Eq. 36}$$

Another drawback that the CLS suffers, is that this latter is sensible to outliers.[28] When a datapoint is out of range of the general trend of the data set, this will badly influence the results of CLS.[28] The example given in reference [28] using the following couples of data ((0, 0.0); (1, 1.1); (2, 2.0); (3, 3.1); (4, 3.8); (5, 10.0)) will give an idea of the problem (Figure 16). The resistance (or breakdown bound) is zero for CLS. The use of the median (which is the central value of an ordered data set) is exceptionally resistant to outliers.[28] In 1950, Theil proposed a resistant and easily implementable method which consists of collecting the slopes between all possible pairs (Equation 37) and find the median of these slopes.[28]

$$b_{ij}=\frac{y_j-y_i}{x_j-x_i} \qquad \text{with } x_i\neq x_j,\ 1\leq i<j\leq n \qquad \text{Eq. 37}$$

This givse n(n-1)/2 slope values. The slope b is given by finding the median of all the above slopes and the intercept a by finding the median of all $\{y_i - b\ x_i\}$. The breakdown bound is 29 %.[28] An improvement of this resistance to 50 % is due to Siegel in 1982. He proposed to calculate the slope b in two steps, first the median of all the slopes calculated from a given datapoint and then the median of all the medians collected as explained in Equation 38. The intercept a is calculated with the same equation as above. This method received Theil-Siegel's name.

$$b = \underset{i}{\operatorname{med}}\{\underset{j \neq i}{\operatorname{med}}\{b_{ij}\}\} \qquad \text{Eq. 38}$$

The results with the above example are for CLS y = -0.895 + 1.691 x (r = 0.894), y = 1.033 x for the Theil method and y = 0.025 + 1.017 x for the Thiel-Siegel method and are shown in Figure 16.

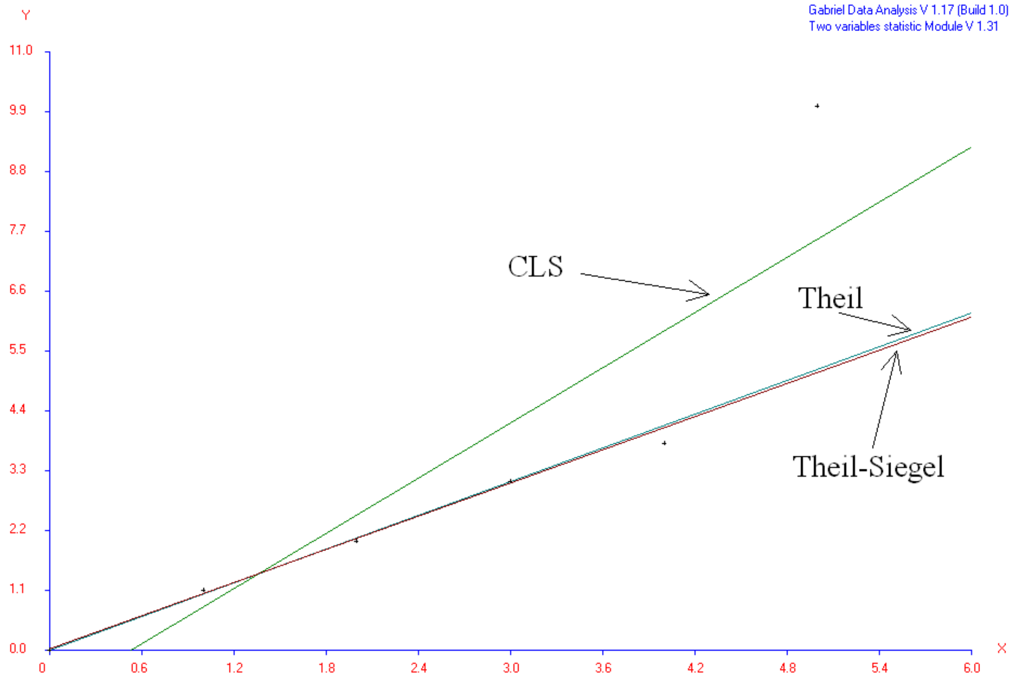


Fig. 16 : Graphical comparison of CLS, Theil and Theil-Siegel regressions.

Practically the Theil and the Theil-Siegel are programmed in GDA as follows. First, all data is ordered by x ascending. Then GDA will search duplicate values of x: if some are found, the calculations will not be performed to prevent a crash with a division by zero. As stated above, the number of slopes is equal to n(n-1)/2 and to avoid an increase of the memory used and a risk of a crash due to missing memory, the number of data in the Theil method is limited to 256 ; if more than 256 points are present, the Theil method is not performed which is not the case of the Theil-Siegel method. The rest of the method is conform of the theoretical method described above.

Another method that has been implemented in GDA is a linear least square regression assuming a constant percent error.[27] In the CLS, it is assumed that the absolute error is constant, which means that the larger values of y influence too much the linear regression. The constant percent error method assumes that the percentage error is constant by minimizing the weighted squares of the deviation from the least squares lines.[27] With this method of analyze, the results between CLS and the constant percent error method are very close but the latter gives a closer fit to the theoretical linear regression. According to reference [27], equations 39 and 40 are implemented in GDA but the calculation is only performed if no y's value is equal to zero.

$$a = \frac{\left(\sum_{i=1}^{n}\frac{1}{y_i}\right)\left(\sum_{i=1}^{n}\frac{x_i^2}{y_i^2}\right) - \left(\sum_{i=1}^{n}\frac{x_i}{y_i^2}\right)\left(\sum_{i=1}^{n}\frac{x_i}{y_i}\right)}{\left(\sum_{i=1}^{n}\frac{1}{y_i^2}\right)\left(\sum_{i=1}^{n}\frac{x_i^2}{y_i^2}\right) - \left(\sum_{i=1}^{n}\frac{x_i}{y_i^2}\right)^2} \qquad \text{Eq. 39}$$

$$b = \frac{\left(\sum_{i=1}^{n}\frac{x_i}{y_i}\right)\left(\sum_{i=1}^{n}\frac{1}{y_i^2}\right) - \left(\sum_{i=1}^{n}\frac{x_i}{y_i^2}\right)\left(\sum_{i=1}^{n}\frac{1}{y_i}\right)}{\left(\sum_{i=1}^{n}\frac{x_i^2}{y_i^2}\right)\left(\sum_{i=1}^{n}\frac{1}{y_i^2}\right) - \left(\sum_{i=1}^{n}\frac{x_i}{y_i^2}\right)^2} \qquad \text{Eq. 40}$$

An explained variable is a variable on which the distance parallel to its axis is minimized.[24] As stated above for the OLS paragraph, CLS minimizes only the squared distance parallel to the y axis. When x is the explained variable, it is equivalent to a change of the axis x and y and the horizontal distances are used instead of the vertical ones. So the definition of x and y are inverted and equations 41 to 44 are those in use in GDA.[24] GDA gives $y = a + b\ x$ and $x = b_0 + b_1\ y$ equations

$$b_0 = \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i^2 - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i\ y_i}{n \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2} \quad \text{Eq. 41 and} \quad b_1 = \frac{n \sum_{i=1}^{n} x_i\ y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2} \quad \text{Eq. 42}$$

and if b_1 is different from zero, then

$$a = \frac{-b_0}{b_1} \quad \text{Eq. 43 and} \quad b = \frac{1}{b_1} \quad \text{Eq. 44}$$

Now the non linear regressions will be discussed here. At page 15 of reference [25] it is explained how to linearize an equation and after, the linearized form is fitted with a least squares method. Table 1 gives the names, the equations, which variable has been linearized and how. The last column needs more explanation. The CLS is used several times in GDA and has been implemented as a subroutine. Sometimes the calculations are performed by the internal CLS subroutine (after linearization) and sometimes the coefficients' equations are programmed directly (coming from historical programming of GDA). Naturally if a datapoint is not mathematically allowed (example the logarithm of a negative value) the corresponding equation is not calculated.

Table 1 : Non linear regressions

Name of the equation or function	**Equation**	**Variable(s) linearized**		**Use the internal CLS subroutines**
		Initial variables	Variables used in calculations	
Logarithmic	$y = a + b \ln x$ Eq. 45	x y	ln x y	No
Power law	$y = a\ x^b$ Eq. 46	x y	ln x ln y	No
Exponential 1[12]	$y = a\ b^x$ Eq. 47	x y	x ln y	No
Exponential 2	$y = e^{(a + b\ x)}$ Eq. 48	x y	x ln y	Yes
Exponential 3[13]	$y = a\ e^{b\ x}$ Eq. 49	x y	x ln y	Yes

[12] a and b are calculated by taking the exponential of the intercept and the slope values obtained by the classical equations.

[13] a is calculated by taking the exponential of the intercept value sent by the internal CLS routine.

Reciprocal	$y=\frac{x}{(a+b\,x)}$ Eq. 50	x y	$\frac{1}{x}$ $\frac{1}{y}$	Yes
Casson[14]	$\sqrt{y}=\sqrt{a}+\sqrt{b\,x}$ Eq. 51	x y	$\sqrt{x}$ $\sqrt{y}$	Yes

The Herschel-Bulkley model is a rheological model using a three-parameter power law equation $y = a + b\,x^n$ (Equation 52). The determination of the three parameters is not easy and most of the time a is first graphically estimated and next, the other parameters, b and n, are determined linearly by using ln(y - a) instead of y. b and n are very sensitive to the estimation of a.[29] In 2008, Mullineux[29] proposed a particular method for resolving the equation 52. Contrary to what is usually carried out, n is found first and then the coefficients a and b by CLS, using x^n instead of x. n is determined when the determinant of the matrix[15] presented above is equal to zero (Equation 53).[29]

$$F(n)=\begin{vmatrix} m & \sum x^n & \sum x^n \ln x \\ \sum x^n & \sum x^{2n} & \sum x^{2n} \ln x \\ \sum y & \sum x^n\,y & \sum x^n\,y \ln x \end{vmatrix}=0 \qquad \text{Eq. } 53^{16}$$

In GDA, n is searched by dichotomy for an n positive value and for an n negative value and then, the value which gives the best correlation coefficient (in the internal CLS subroutine with x^n) is used for calculated the a and b coefficients.

The next part of this module is dedicated to polynomials and the first, called Linear polynomial 2 ($y = a + b\,x^2$ (Equation 54)), is simply programmed by using the internal CLS subroutine after changing the variable x in x power two.

Now the part of this module is devoted to polynomials of the form $y = a_0 + a_1\,x + a_2\,x^2 + a_3\,x^3 + ... + a_n\,x^n$ (Equation 55).[30] GDA searches nth degree polynomial equations with n from 2 up to 6.[30] It seems to us that higher degrees are not necessary as most of the time, the coefficients for the highest degrees are very little and correspond to little corrections (see Figure 15). The determination of the equations for this nth degree polynomials uses classical

[14] a and b are calculated by taking the square of the intercept and the slope values returned by the internal CLS routine.

[15] Σ means $\sum\limits_{i=1}^{m}$

resolution.[23, 24] First GDA will construct some matrices. The first one corresponds to the theoretical matrix presented as equation 56 (Id is the order of the polynomial wanted, n the total number of points and $\sum$ means $\sum\limits_{i=1}^{n}$).

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 & \dots\dots\dots & \sum x_i^{Id} \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \dots\dots\dots & \sum x_i^{Id+1} \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \dots\dots\dots & \sum x_i^{Id+2} \\ & & & & \\ \sum x_i^{Id} & \sum x_i^{Id+1} & \sum x_i^{Id+2} & \dots\dots\dots & \sum x_i^{2\,Id} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \\ a_{Id} \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum y_i\, x_i \\ \sum y_i\, x_i^2 \\ \\ \sum y_i\, x_i^{Id} \end{bmatrix} \qquad \text{Eq. 56}$$

In fact GDA uses a slightly modified matrix called r (equation 57).

$$r = \begin{bmatrix} n & \sum x_i & \sum x_i^2 & \dots\dots\dots & \sum x_i^{Id} & \sum y_i \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \dots\dots\dots & \sum x_i^{Id+1} & \sum y_i\, x_i \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \dots\dots\dots & \sum x_i^{Id+2} & \sum y_i\, x_i^2 \\ & & & & & \\ \sum x_i^{Id} & \sum x_i^{Id+1} & \sum x_i^{Id+2} & \dots\dots\dots & \sum x_i^{2\,Id} & \sum y_i\, x_i^{Id} \end{bmatrix} \qquad \text{Eq. 57}$$

Then, the different coefficients a_x are determined by the resolution of the system by the method of Gauss-Jordan.[23] The different coefficients are read in the last column of the matrix. As the matrix r is modified during the resolution by the method of Gauss-Jordan, two other matrices are constructed, called a' and t which are presented in equations 58 and 59.

$$a' = \begin{bmatrix} n \\ \sum x_i \\ \sum x_i^2 \\ \\ \sum x_i^{2\,Id} \end{bmatrix} \qquad \text{Eq. 58} \qquad\qquad t = \begin{bmatrix} \sum y_i \\ \sum y_i\, x_i \\ \sum y_i\, x_i^2 \\ \\ \sum y_i\, x_i^{Id} \\ \sum y_i^2 \end{bmatrix} \qquad \text{Eq. 59}$$

The correlation coefficient is calculated according to equations 60 to 62 using two intermediate variables P and q.

[16] Exceptionally m means the total number of data instead of the usually n.

$$P = \sum_{j=1}^{Id} a_j \left(\sum_{i=1}^{n} y_i \; x_i^j - \sum_{i=1}^{n} x_i^j \frac{\sum_{i=1}^{n} y_i}{n} \right) \quad \text{Eq. 60} \qquad \text{with} \qquad q = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i \right)^2}{n} \quad \text{Eq. 61}$$

$$r = \sqrt{\frac{P}{q}} \qquad \text{Eq. 62}$$

The last piece of information that the user can find is just below the correlation coefficient and is the standard error of the estimation (SEE) (Equation 63).

$$SEE = \sqrt{\frac{q - P}{n - Id - 1}} \qquad \text{Eq. 63}$$

The last equation is a three parameter equation and is the hyperbolic function (Equation 64).[25] It is resolved by solving, by the method of Gauss-Jordan[23], the equation's system constituted by equations 65 to 67 (n is the total number of points and $\sum$ means $\sum_{i=1}^{n}$).[25]

$$y = \frac{(a + b\,x)}{(1 + c\,x)} \qquad \text{Eq. 64}$$

$$\sum y_i = a\,n + b \sum x_i - c \sum x_i \; y_i \qquad \text{Eq. 65}$$

$$\sum x_i \; y_i = a \sum x_i + b \sum x_i^2 - c \sum x_i^2 \; y_i \qquad \text{Eq. 66}$$

$$\sum x_i \; y_i^2 = a \sum x_i \; y_i + b \sum x_i^2 \; y_i - c \sum x_i^2 \; y_i^2 \qquad \text{Eq. 67}$$

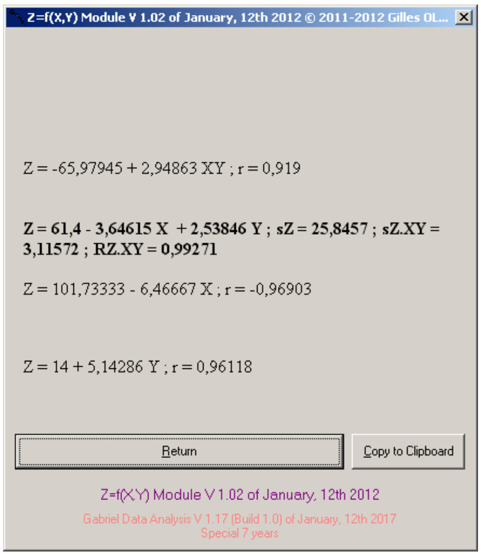
In fact GDA uses a matrix that looks like the matrix r of equation 57. To have a idea of the quality of the fit, the standard deviation S is given and follows the equation 68, in which y_i is the experimental value and y_c is the calculated value for the same x_i.[25]

$$S = \sqrt{\frac{\sum (y_i - y_c)^2}{n - 3}} \qquad \text{Eq. 68}$$

Naturally, if new equations are found, then they will be introduced in this module, one of the big reasons of GDA. This module, thanks to the panel where all equations are written in one time, is easier to use then Microsoft Excel.

4.4. Z = f(X, Y) Module (actual version: 1.02)

This module is the beginning of experimental design. In this way, this module uses a subroutine that was implemented in prevision of a bigger module devoted to experimental design. In this module four calculations are carried out. The first, the third and the fourth use the internal CLS subroutine and will not be discussed here, only the first uses a change of variable, sending to the routine the product of X and Y instead of a monovariable (Figure 17). The second gives the regression plane's equation z = a + b x + c y (Equation 69).



Z=f(X,Y) Module V 1.02 of January, 12th 2012 © 2011-2012 Gilles OL...

```
Z = -65,97945 + 2,94863 XY ; r = 0,919

Z = 61,4 - 3,64615 X  + 2,53846 Y ; sZ = 25,8457 ; sZ.XY =
3,11572 ; RZ.XY = 0,99271

Z = 101,73333 - 6,46667 X ; r = -0,96903

Z = 14 + 5,14286 Y ; r = 0,96118
```

Return
Copy to Clipboard
Z=f(X,Y) Module V 1.02 of January, 12th 2012
Gabriel Data Analysis V 1.17 (Build 1.0) of January, 12th 2017
Special 7 years

Fig. 17 : The Z = f(X,Y) Module screen result.

As already mentioned, this module uses the same subroutine as the next module (Experimental Design) and the algorithm will be discussed here. This one uses a matrix, called RetourPlanExp (Equation 70)[24], similar to the matrix r to find polynomials' equations (Equation 57). In Equation 70, Id is the number of variables, n the total number of points,

$\sum X_A$ means $\sum_{i=1}^{n} x_{A_i}$ and in the case of this module Z = f(X,Y), X_1 is the value in column X, X_2 is the value in column Y, and Y is the value in column Z in Grid1 from the Input Module.

$$\text{RetourPlanExp} = \begin{bmatrix} n & \sum X_1 & \sum X_2 & \ldots\ldots\ldots\ldots\ldots\ldots & \sum X_{Id} & \sum y_i \\ \sum X_1 & \sum X_1^2 & \sum X_2 X_1 & \ldots\ldots\ldots\ldots\ldots\ldots & \sum X_{Id} X_1 & \sum y_i\, X_1 \\ \sum X_2 & \sum X_1 X_2 & \sum X_2^2 & \ldots\ldots\ldots\ldots\ldots\ldots & \sum X_{Id} X_2 & \sum y_i\, X_2 \\ & & & & & \\ \sum X_{Id} & \sum X_1 X_{Id} & \sum X_2 X_{Id} & \ldots\ldots\ldots\ldots\ldots\ldots & \sum X_{Id}^2 & \sum y_i\, X_{Id} \end{bmatrix} \qquad \text{Eq. 70}$$

The different coefficients $a_x{}^{17}$ are found by the resolution of the matrix RetourPlanExp by the method of Gauss-Jordan[23] and then are read in the last column of the matrix. To have an idea of the accuracy of the fit, three parameters are calculated.[24] The first one is the standard deviation of the z values alone and is classically defined as equation 71 and depend only of the experimental data.

$$s_Z = \sqrt{\frac{\sum_{i=1}^{n} z_i^2}{n} - \left(\frac{\sum_{i=1}^{n} z_i}{n}\right)^2} \qquad \text{Eq. 71}$$

The tied up standard deviation $s_{Z.XY}$ of the estimator of z obtained from x and y is depending on the experimental data and on the estimated values of z (Equation 72).

$$s_{Z.XY} = \sqrt{\frac{\sum_{i=1}^{n} \left(z_i - z_{i_{estimated}}\right)^2}{n}} \qquad \text{Eq. 72}$$

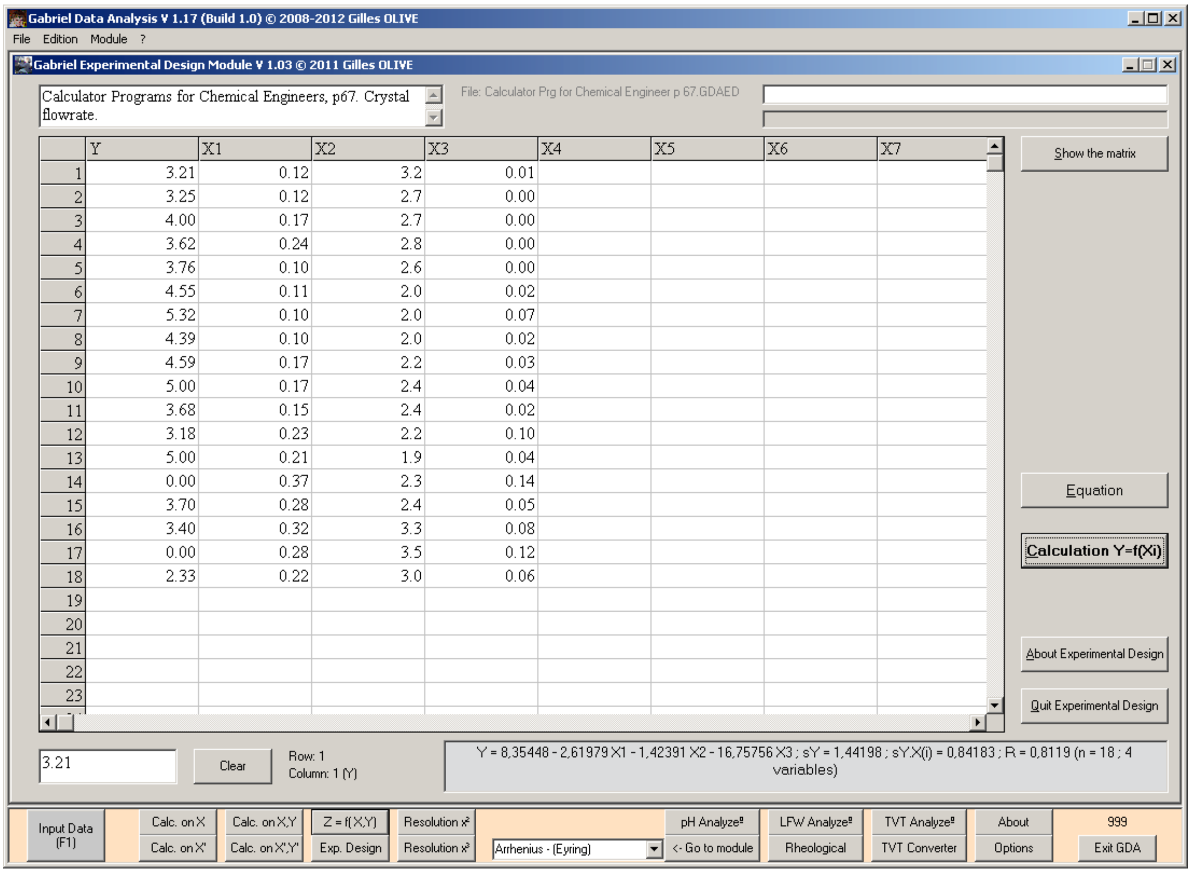
At least the multiple correlation coefficient looks like Equation 36 for the linear correlation coefficient and is given in equation 73.

$$R_{Z.XY} = \sqrt{1 - \frac{s_{Z.XY}^2}{s_Z^2}} \qquad \text{Eq. 73}$$

4.5. Experimental Design Module (actual version: 1.03)

This editor comes with the same functions as the Input Module and with an TXT Import Module (actual version: 1.00). Figure 18 shows a typical screen. Up to ten independent variables can be entered[18]. The results can be selected and pasted into any word processing. The matrix before and after the inversion can be seen in a supplementary window (Figure 19). Two types of calculations are possible.

The first one is resolution of $y = f(x_i)$. This section will be short as the algorithm is the same as in the section 4.4 above (Z = f(X, Y) Module) and the reader can refer to equations 70 to 73. Unfortunately the trivial cases are not always found: this comes from the method of Gauss-Jordan which is very sensitive to the number of digits for the calculations.



Gabriel Data Analysis V 1.17 (Build 1.0) © 2008-2012 Gilles OLIVE

Gabriel Experimental Design Module V 1.03 © 2011 Gilles OLIVE

Calculator Programs for Chemical Engineers, p67. Crystal flowrate.

File: Calculator Prg for Chemical Engineer p 67.GDAED

	Y	X1	X2	X3	X4	X5	X6	X7
1	3.21	0.12	3.2	0.01				
2	3.25	0.12	2.7	0.00				
3	4.00	0.17	2.7	0.00				
4	3.62	0.24	2.8	0.00				
5	3.76	0.10	2.6	0.00				
6	4.55	0.11	2.0	0.02				
7	5.32	0.10	2.0	0.07				
8	4.39	0.10	2.0	0.02				
9	4.59	0.17	2.2	0.03				
10	5.00	0.17	2.4	0.04				
11	3.68	0.15	2.4	0.02				
12	3.18	0.23	2.2	0.10				
13	5.00	0.21	1.9	0.04				
14	0.00	0.37	2.3	0.14				
15	3.70	0.28	2.4	0.05				
16	3.40	0.32	3.3	0.08				
17	0.00	0.28	3.5	0.12				
18	2.33	0.22	3.0	0.06				
19								
20								
21								
22								
23								

3.21

Row: 1
Column: 1 (Y)

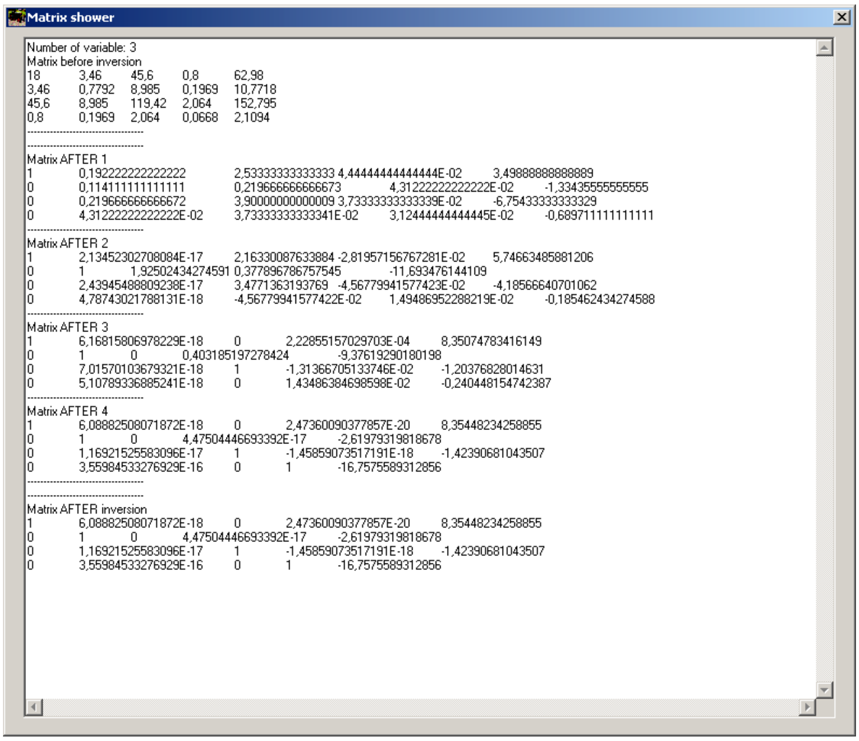
Y = 8,35448 - 2,61979 X1 - 1,42391 X2 - 16,75756 X3 ; sY = 1,44198 ; sY.X(i) = 0,84183 ; R = 0,8119 (n = 18 ; 4 variables)

Arrhenius - (Eyring)

Fig. 18 : Experimental Design Module screen.

[17] for the hyperplane $y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + ... + a_n x_n$.

The second type of calculation is the resolution of n equations with n unknowns.[23] In this case the grid is read and the matrix corresponds to the grid itself and those matrices will be resolved again by the method of Gauss-Jordan.[23]



```
Number of variable: 3
Matrix before inversion
18        3,46      45,6      0,8       62,98
3,46      0,7792    8,985     0,1969    10,7718
45,6      8,985     119,42    2,064     152,795
0,8       0,1969    2,064     0,0668    2,1094
----------------------------------

----------------------------------
Matrix AFTER 1
1         0,192222222222222         2,53333333333333 4,44444444444444E-02         3,49888888888889
0         0,114111111111111         0,219666666666673         4,31222222222222E-02         -1,33435555555555
0         0,219666666666672         3,90000000000009 3,73333333333339E-02         -6,75433333333329
0         4,31222222222222E-02      3,73333333333341E-02      3,12444444444445E-02         -0,689711111111111
----------------------------------
Matrix AFTER 2
1         2,13452302708084E-17      2,16330087633884 -2,81957156767281E-02        5,74663485881206
0         1         1,92502434274591 0,377896786757545         -11,693476144109
0         2,43945488809238E-17      3,4771363193769  -4,56779941577423E-02        -4,18566640701062
0         4,78743021788131E-18      -4,56779941577422E-02     1,49486952288219E-02         -0,18546243427458
----------------------------------
Matrix AFTER 3
1         6,16815806978229E-18      0         2,22855157029703E-04      8,35074783416149
0         1         0         0,403185197278424         -9,37619290180198
0         7,01570103679321E-18      1         -1,31366705133746E-02     -1,20376828014631
0         5,10789336885241E-18      0         1,43486384698598E-02      -0,240448154742387
----------------------------------
Matrix AFTER 4
1         6,08882508071872E-18      0         2,47360090377857E-20      8,35448234258855
0         1         0         4,47504446693392E-17      -2,61979319818678
0         1,16921525583096E-17      1         -1,45859073517191E-18     -1,42390681043507
0         3,55984533276929E-16      0         1         -16,7575589312856
----------------------------------

----------------------------------
Matrix AFTER inversion
1         6,08882508071872E-18      0         2,47360090377857E-20      8,35448234258855
0         1         0         4,47504446693392E-17      -2,61979319818678
0         1,16921525583096E-17      1         -1,45859073517191E-18     -1,42390681043507
0         3,55984533276929E-16      0         1         -16,7575589312856
```

Fig. 19 : Matrix shower window during resolution by the method of Gauss-Jordan.

4.6. Equation Resolver Module (actual version: 1.04)

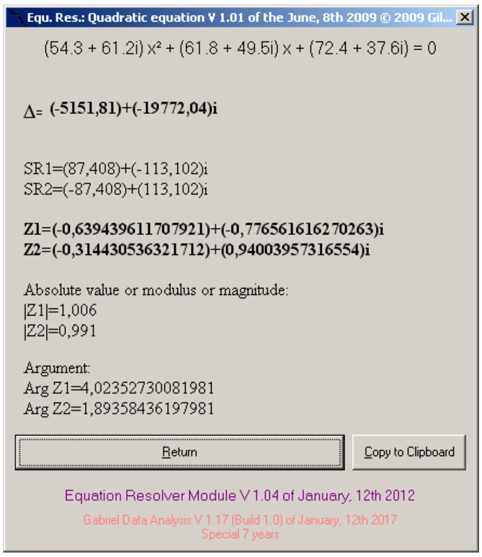
This module is not a big one and must be seen as a tool. GDA is able to solve quadratic equations by pressing the Resolution x² button (Figure 20) and cubic equation by pressing Resolution x³ (Figure 21).

4.6.1. Quadratic Equation (actual version: 1.01)

This module will resolve equations of the form a $x^2 + b\ x + c = 0$ (Eq. 74). GDA is able to resolve equation 74 with a, b and c coefficients, real or imaginary. Cells X1, X2 and X3 are used for introducing respectively a, b and c if they are real and Y1, Y2 and Y3 for the

[18] It is possible to have more than ten variables if a user asks for it.

imaginary part with X1, X2 and X3 for the real part. Naturally GDA is able to give imaginary results also if the coefficients are real only. Figure 20 shows the result screen.


Equ. Res.: Quadratic equation V 1.01 of the June, 8th 2009 © 2009 Gil...
Return
Copy to Clipboard

```
(54.3 + 61.2i) x² + (61.8 + 49.5i) x + (72.4 + 37.6i) = 0

Δ= (-5151,81)+(-19772,04)i

SR1=(87,408)+(-113,102)i
SR2=(-87,408)+(113,102)i

Z1=(-0,639439611707921)+(-0,776561616270263)i
Z2=(-0,314430536321712)+(0,94003957316554)i

Absolute value or modulus or magnitude:
|Z1|=1,006
|Z2|=0,991

Argument:
Arg Z1=4,02352730081981
Arg Z2=1,89358436197981
```

Equation Resolver Module V 1.04 of January, 12th 2012

Gabriel Data Analysis V 1.17 (Build 1.0) of January, 12th 2017
Special 7 years

Fig. 20 : Result screen of Quadratic Equation (in Equation Resolver Module) of GDA.

In the following text the subscript r means the real part of a coefficient while the subscript i means the imaginary part.

If the coefficient a, b and c are real, the first calculation is the calculation of the discriminant Δ (equation 75).

$$\Delta = b_r^2 - 4\, a_r\, c_r \qquad \text{Eq. 75}$$

At this point three cases exist depending of the value of Δ. If Δ is equal to zero then there is a unique solution (equation 76) while if Δ is positive then the system admits two solutions X1 and X2 according to equation 77.

$$X1 = X2 = \frac{-\,b_r}{2\,a_r} \qquad \text{Eq. 76}$$

$$X1 = \frac{-\,b_r - \sqrt{\Delta}}{2\,a_r} \text{ and } X2 = \frac{-\,b_r + \sqrt{\Delta}}{2\,a_r} \qquad \text{Eq. 77}$$

The two first solutions are trivial and well-known. The last case is the one where Δ is negative. At this point the treatment is the same if the user has introduced imaginary coefficients, also if they are all equal to zero, and in case of a negative Δ, the imaginary parts of the coefficients are set to zero. The real part and the imaginary part must be calculated separately. So the discriminant Δ become (equation 78)

$$\Delta_r = b_r^2 - b_i^2 - 4\,a_r\,c_r + 4\,a_i\,c_i \text{ and } \Delta_i = 2\,b_r\,b_i - 4\,c_r\,a_i - 4\,c_i\,a_r \text{ then } \Delta = \Delta_r + \Delta_i i \quad \text{Eq. 78}$$

The two solutions, called Z1 and Z2, are found by equation 84 but before equations 79 and 80 define the square root of the discriminant[31] and equations 81 to 83 intermediate variables.

$$SR_r = \sqrt{\frac{\left|\sqrt{\Delta_r^2 + \Delta_i^2} + \Delta_r\right|}{2}} \text{ and } SR_i = \sqrt{\frac{\left|\sqrt{\Delta_r^2 + \Delta_i^2} - \Delta_r\right|}{2}} \quad \text{Eq. 79}$$

$$SR1 = SR_r + SR_i i \text{ and } SR2 = -SR_r - SR_i i \quad \text{Eq. 80}$$

The sign of R_i is changed if 2 R_r R_i is different from Δ_i.

$$S_{1r} = 2\,a_r\,(-b_r + R_r) + 2\,a_i\,(R_i - b_i) \text{ and } S_{1i} = 2\,a_r\,(-b_i + R_i) - 2\,a_i\,(R_r - b_r) \quad \text{Eq. 81}$$

$$S_{2r} = 2\,a_r\,(-b_r - R_r) + 2\,a_i\,(-R_i - b_i) \text{ and } S_{2i} = 2\,a_r\,(-b_i - R_i) - 2\,a_i\,(-R_r - b_r) \quad \text{Eq. 82}$$

$$P = 4\,a_r^2\,a_i^2 \quad \text{Eq. 83}$$

$$Z1 = \frac{S_{1r}}{P} + \frac{S_{1i}}{P} i \text{ and } Z2 = \frac{S_{2r}}{P} + \frac{S_{2i}}{P} i \quad \text{Eq. 84}$$

The following equations (equations 85 and 86) give other information in particular the modulus or magnitude of Z, which can be considered as the "absolute value" of the imaginary number and it is a positive or a null real, and the argument (arg Z) which is as the angle between the x axis and the line from the point 0,0 and the image of the imaginary number in a complex plane.[31, 32] The argument is a real number.[31-33]

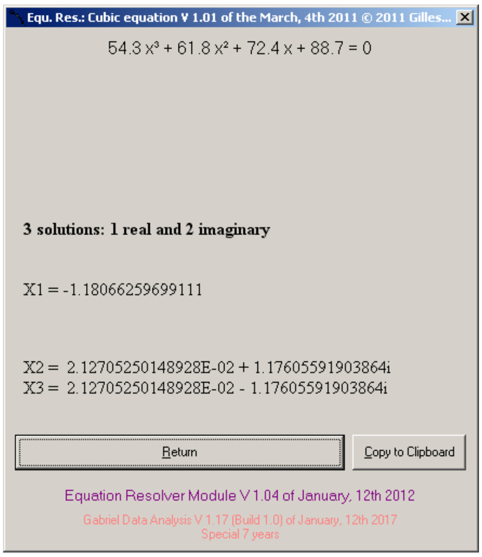
$$|Z1| = \sqrt{\frac{S_{1r}^2 + S_{1i}^2}{P^2}} \text{ and } |Z2| = \sqrt{\frac{S_{2r}^2 + S_{2i}^2}{P^2}} \qquad \text{Eq. 85}$$

$$\text{Arg } Z1 = \tan^{-1}\left(\frac{S_{1i}}{S_{1r}}\right) \text{ and } \text{Arg } Z2 = \tan^{-1}\left(\frac{S_{2i}}{S_{2r}}\right) \qquad \text{Eq. 86}$$

Naturally equation 86 is calculated only if $S1_r$ and $S2_r$ are different of zero. If $S1_r$ or $S2_r$ is negative then Pi is added to the corresponding result.

4.6.1. Cubic Equation (actual version: 1.01)

$a\ x^3 + b\ x^2 + c\ x + d = 0$ (Eq. 87) is the form that computes GDA. GDA is able to resolve equation 87 with a, b, c and d coefficients that are real only. In this case cells X1, X2, X3 and X4 are used for introducing respectively a, b, c and d. Figure 21 shows a typical result screen.



```
Equ. Res.: Cubic equation V 1.01 of the March, 4th 2011 © 2011 Gilles...

54.3 x³ + 61.8 x² + 72.4 x + 88.7 = 0

3 solutions: 1 real and 2 imaginary

X1 = -1.18066259699111

X2 =  2.12705250148928E-02 + 1.17605591903864i
X3 =  2.12705250148928E-02 - 1.17605591903864i
```

Return
Copy to Clipboard
Equation Resolver Module V 1.04 of January, 12th 2012
Gabriel Data Analysis V 1.17 (Build 1.0) of January, 12th 2017
Special 7 years

Fig. 21 : Result screen of Cubic Equation (in Equation Resolver Module) of GDA.

The resolution of equation 87, a cubic equation, is more complicated than the resolution of a quadratic equation. The method used here is a slightly modified method of the Marseiller Bernard Sotta and is described as algorithm in the supporting information.[34] Schematically, GDA searches the trivial triple and double solutions and otherwise the three solutions by dichotomy. In this latter, if needed, complex solutions are given (Figure 21).

4.7. Arrhenius/Eyring Module (actual version: 1.00)

This module is accessible only by the combobox list.

In 1889, Arrhenius, working on reaction's kinetic, found experimentally a relation between the reaction rate k and the temperature which is presented in equation 88 (A is the pre-exponential factor or the "action constant" (A), E_a the energy of activation , R is the gas constant (8.314472 $J.mol^{-1}.K^{-1}$) and T the absolute temperature in Kelvin ($T_K = t_{°C} + 273.15$)).

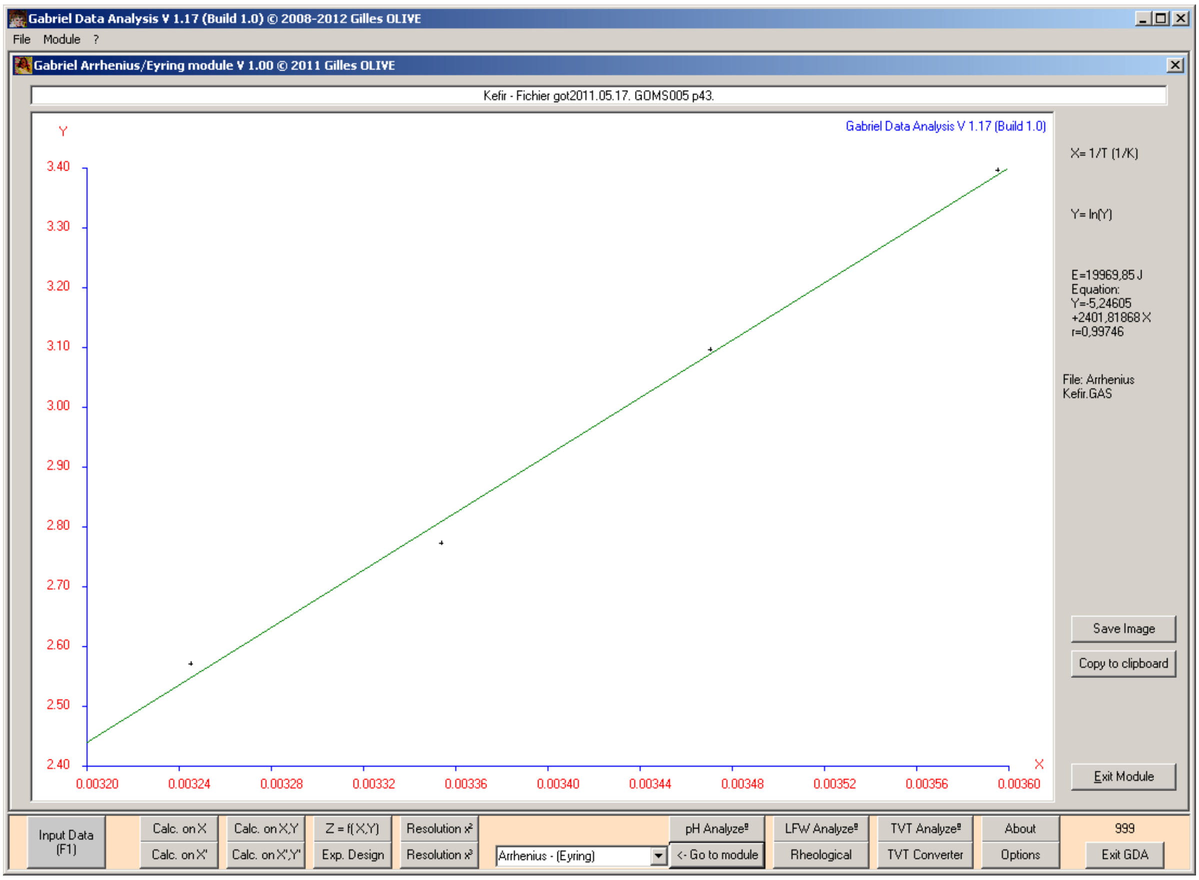
$$k = A.e^{\left(\frac{-E_a}{R.T}\right)} \qquad \text{Eq. 88}$$

Later, the equation (88) was theoretically explained by Eyring in 1935. As this latter also worked on rheology, equation 88 became equation 89 (η_a is the apparent viscosity (Pa.s), η_0 is considered as the viscosity at infinite temperature (mPa.s) and E_a the (standard free) energy of activation (by mole) written sometimes ΔE_{visco}[19]. The other constants have the same meaning as in equation 88).[35]

$$\eta_a = \eta_0.e^{\left(\frac{E_a}{R.T}\right)} \qquad \text{Eq. 89}$$

In the two cases, the naperian logarithm as a function of the inverse of the absolute temperature is plotted to find the parameters. In GDA, this is equation 89 that has been implemented as can be seen in Figure 22. The values are found by the internal CLS subroutine using (1/(X+273.15)) as x and (ln Y) as y, the slope being equal to E/R and the intercept to η_0 (see Figure 22 in the right part).

[19] It seems interesting to compare the activation energy of viscosity and the variation of the internal energy of vaporization (study of Ewell and Eyring). Kincaid gives the empirical law $\Delta E_{visco} = \Delta U_{vap} / 2.45$ (not implemented in GDA).



Gabriel Data Analysis V 1.17 (Build 1.0) © 2008-2012 Gilles OLIVE

Gabriel Arrhenius/Eyring module V 1.00 © 2011 Gilles OLIVE

Kefir - Fichier got2011.05.17. GOMS005 p43.

X= 1/T (1/K)

Y= ln(Y)

E=19969,85 J
Equation:
Y=-5,24605
+2401,81868 X
r=0,99746

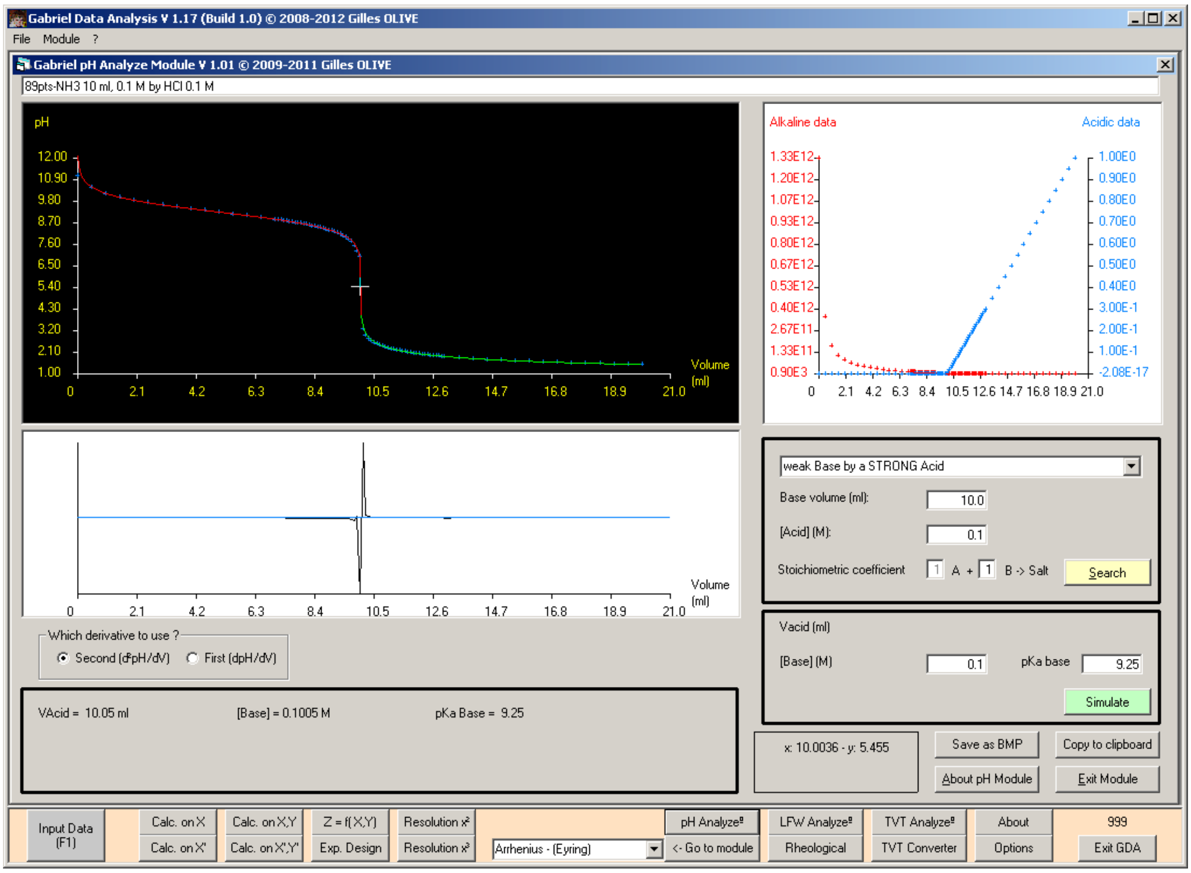
File: Arrhenius
Kefir.GAS

Arrhenius - (Eyring)

Fig. 22 : Arrhenius-Eyring module's results.

4.8. pH Analyze Module (actual version: 1.01)

The aim of this module is more for education than for research. After entering the datapoints and set up the flag to pH in the Input Module, Figure 23 shows a typical screen. This module can either simulate a curve (Simulate button) or calculate with the data the missing information (Search button).



Gabriel Data Analysis V 1.17 (Build 1.0) © 2008-2012 Gilles OLIVE

Gabriel pH Analyze Module V 1.01 © 2009-2011 Gilles OLIVE

89pts-NH3 10 ml, 0.1 M by HCl 0.1 M

weak Base by a STRONG Acid

Base volume (ml): 10.0

[Acid] (M): 0.1

Stoichiometric coefficient 1 A + 1 B -> Salt

Vacid (ml)

[Base] (M) 0.1

pKa base 9.25

Which derivative to use ?

● Second (d²pH/dV) ○ First (dpH/dV)

VAcid = 10.05 ml

[Base] = 0.1005 M

pKa Base = 9.25

x: 10.0036 - y: 5.455

Arrhenius - (Eyring)

Fig. 23 : pH Analyze Module screen.

The different equations will not be discussed here and the reader will find them in the supporting information.

In the screen from the top to bottom, first the number of points is found with the information from the Input Module's comments textbox, followed below on the left by the plot of the pH as a function of the titrant added, the crosses are the experimental points and the red and green lines the theoritical curve of which the constants are determined by experimental data (Search button) or by the constants given by the user (Simulate button). More below, the plot of the derivative as a function of the titrant added ; in blue the line equal to zero. The derivative, first[20] or second[21], used for determining the equivalent point can be chosen, the second derivative giving better results, but sometimes cannot work. The results are indicated in the square: equivalent volume, concentration of the titrant and if this is the case, the pKa. In the right part of the screen, in the top, the Gran plot curves (the equations

[20] In this case GDA takes the volume when the derivative is maximum for calculating the equivalent point.

will be discussed hereafter), then the type of titration (this must be chosen first) and next the experimental data of the titration and if the user wants to simulate an experiment the data of this simulation.

Two cases must be distinguished for the Gran plot.[36] In the first, during an acid titration with a base, the plot for the acidic data is the v 10^{-pH} as a function of v, the volume of base added, while the alkaline plot corresponds to $(v_0+v)10^{pH}$ as a function of v, with v the volume of the titrant and v_0 the sample volume.[36] The second case is the titration of a base by an acid. Here the alkaline data is defined as $(v_0+v)10^{pH}$ as a function of v and the acidic data as $(v_0+v)10^{-pH}$ as a function of v, v and v_0 have the same meaning as above.[36]

4.9. LFW Analyze Module (actual version: 1.17)

The aim of this module is more for research than for education, as it is complex by the number of calculations. This module is used for analyzing the monolayer's isotherm obtained by the Lauda FW 2 Langmuir-Balance. The principle of a Langmuir balance will not be discussed here and can be found elsewhere.[37-39] The screen is laid out in five parts as can be seen in Figure 24. The raw data are plotted in part 1 of the screen. As the Langmuir balance is sensitive to the noise, the first thing to do is to reduce the noise by smoothing the data. The algorithm used here is the same as the one already described in the Input Module on the x and y's values. In the same time, the calculation will take into account whether the molecular weight is known or not: if the molecular is known, the result will be expressed in area by molecule (equation 90), or else in area by mg (equation 91), and part 3 and 5 (with a new variable change: x will be changed to it naperian logarithm) of the screen will be plotted.

$$Y = \frac{\text{Smoothed Experimental Data} \times \text{Molecular Weight}}{\left(\left(\frac{\text{Solute Weight}}{\text{Solution Volume}}\right) \times \text{Spreaded Volume} \times 60230\right)} \qquad \text{Eq. 90}$$

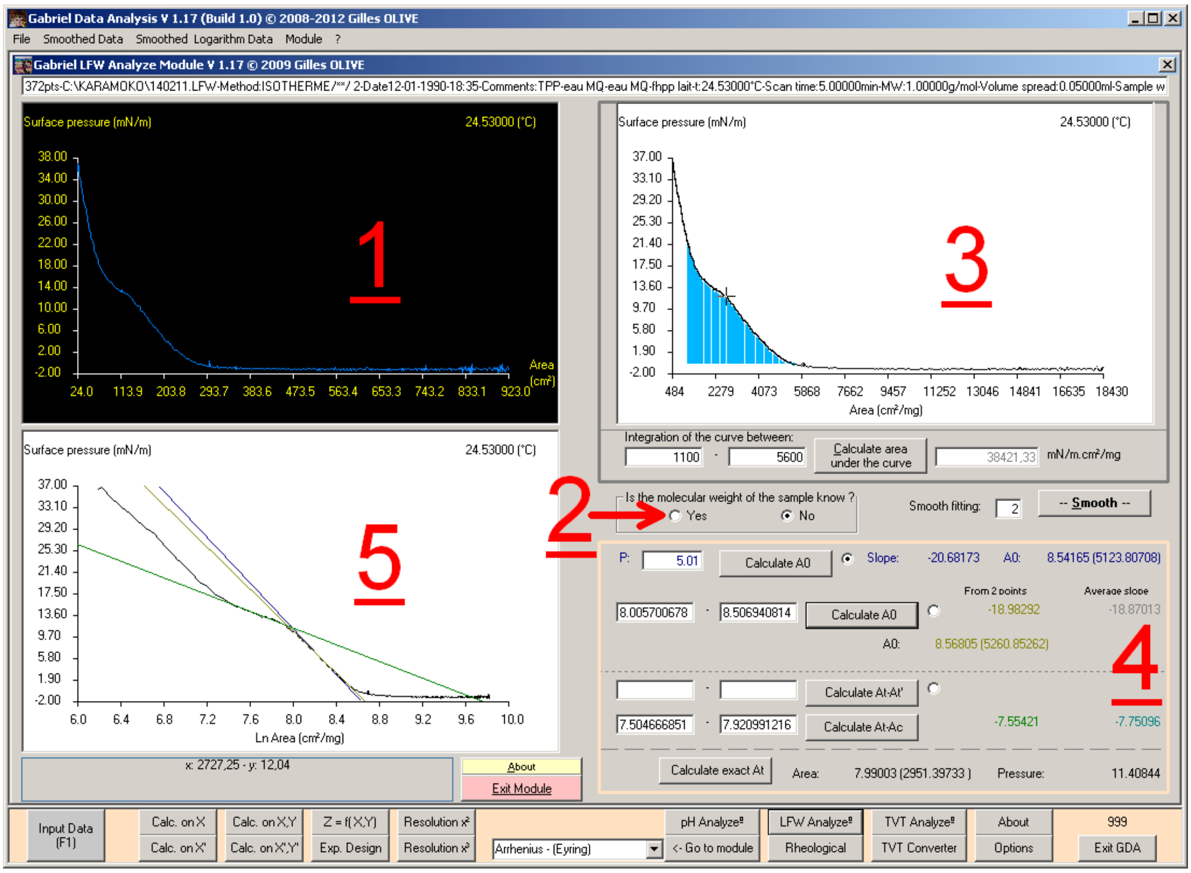
$$Y = \frac{\text{Smoothed Experimental Data}}{\left(\left(\frac{\text{Solute Weight}}{\text{Solution Volume}}\right) \times \text{Spreaded Volume}\right)} \qquad \text{Eq. 91}$$

[21] For the equivalent point, GDA searches, when the second derivative is equal to zero by linear interpolation between the two points of the derivative that change of sign.

The free energy of compression (ΔG_c ($J.g^{-1}$)) defines the work necessary for compressing the film. It corresponds to the surface under the compression isotherm curve and is determined by equation 92 in which π is the surface pressure ($mN.m^{-1}$), dA is the variation of the molecular area ($m^2.mg^{-1}$).[38] A_i and A_f are respectively the surface of the film at the beginning and at the end of compression.

$$\Delta G_c = -\int_{A_f}^{A_i} \pi \, dA \qquad \text{Eq. 92}$$

Part 3 of the screen is devoted to this calculation and uses the already described trapeze's method (Equation 29) between the two indicated boundaries A_i and A_f (text boxes).[23]



372pts-C:\KARAMOKO\140211.LFW-Method:ISOTHERME/**/ 2-Date12-01-1990-18:35-Comments:TPP-eau MQ-eau MQ-fhpp lait-t:24.53000°C-Scan time:5.00000min-MW:1.00000g/mol-Volume spread:0.05000ml-Sample w

Integration of the curve between: 1100 - 5600 38421,33 mN/m.cm²/mg

Is the molecular weight of the sample know ? ○ Yes ● No

Smooth fitting: 2

P: 5.01 ● Slope: -20.68173 A0: 8.54165 (5123.80708)

From 2 points Average slope

8.005700678 - 8.506940814 ○ -18.98292 -18.87013

A0: 8.56805 (5260.85262)

○

7.504666851 - 7.920991216 -7.55421 -7.75096

Area: 7.99003 (2951.39733) Pressure: 11.40844

x: 2727,25 - y: 12,04

Arrhenius - (Eyring)

Fig. 24 : LFW Analyze Module screen.

In part 4 of the screen, the user can calculate the initial area A_0 (which is the surface when no compression is active) or the inflexion At (which could be a change of state or a collapse) and the user can see in part 5 the graphical results. All calculations are performed on the naperian logarithm of x and of y according to equation 90 or 91. The way is always the same for all calculations: the user enters the boundaries x in the textboxes (caution: under logarithm form), GDA searches the nearest datapoint and then performs the calculation(s).

The first point discussed here is the search of A_0. One of the ways is to calculate this value for a pressure of 5 mN.m^{-1}, the slope is calculated first according equation 93 and A_0 from equations 94 and 95.

$$\mathrm{Slope}_{5\,(\mathrm{at\ n})} = \frac{\mathrm{SurfacePressure}_{\mathrm{n\text{-}1}} - \mathrm{SurfacePressure}_{\mathrm{n+1}}}{\mathrm{LnArea}_{\mathrm{n\ \text{-}1}} - \mathrm{LnArea}_{\mathrm{n+1}}} \qquad \text{Eq. 93}$$

$$\mathrm{X}_{5\,(\mathrm{at\ n})} = \mathrm{SurfacePressure}_{\mathrm{n}} - \mathrm{Slope}_{5\,(\mathrm{at\ n})} \times \mathrm{LnArea}_{\mathrm{n}} \qquad \text{Eq. 94}$$

$$\mathrm{A}_{0_{5\,(\mathrm{at\ n})}} = \mathrm{e}^{-\frac{\mathrm{X}_{5\,(\mathrm{at\ n})}}{\mathrm{Slope}_{5\,(\mathrm{at\ n})}}} \qquad \text{Eq. 94}$$

The second way of calculating the slope and A_0 is by introducing the boundaries in the textboxes (under logarithm form). Here two calculations are performed. The first one uses the same equations 93 to 95 in which n-1 is the second boundary while n and n+1 are the first boundary. The second calculation uses the average of each slope (equation 93) calculated between the two boundaries.

In some cases a second transition occurs in the curve (A_t-$A_{t'}$). The algorithm used here is the same as above, for the second way of calculating the slope and A_0, but only the slopes are calculated (equations 93 and 94). This is also the same for the last part of the curve (A_t-A_c).

It is important to find the exact inflexion. Therefore, first the user has to choose the first part (A_{0_5}, $A_{0_{\mathrm{Boundaries}}}$ or A_t-$A_{t'}$) of the curve by selecting the appropriate optionbutton, as the second part is always the curve A_t-A_c. The co-ordinates are found by equations 96 to 98.

$$\mathrm{XX}_{(\mathrm{at\ t})} = \frac{\mathrm{X}_{\mathrm{At\text{-}Ac}_{2\ \mathrm{points}}} - \mathrm{X}_{5,\ \mathrm{Boundaries}_{2\ \mathrm{points}}\ \mathrm{or\ At\text{-}At'}_{2\ \mathrm{points}}}}{\mathrm{Slope}_{5,\ \mathrm{Boundaries}_{2\ \mathrm{points}}\ \mathrm{or\ At\text{-}At'}_{2\ \mathrm{points}}} - \mathrm{Slope}_{\mathrm{At\text{-}Ac}_{2\ \mathrm{points}}}} \qquad \text{Eq. 96}$$

and

$$\mathrm{Area}_{(\mathrm{at\ t})} = \mathrm{e}^{\mathrm{XX}_{(\mathrm{at\ t})}} \qquad \text{Eq. 97}$$

$$\mathrm{Pressure}_{(\mathrm{at\ t})} = \mathrm{X}_{\mathrm{At\text{-}Ac}_{2\ \mathrm{points}}} + \mathrm{Slope}_{\mathrm{At\text{-}Ac}_{2\ \mathrm{points}}} \times \mathrm{XX}_{(\mathrm{at\ t})} \qquad \text{Eq. 98}$$

It could be useful to plot the compressibility or the elasticity as a function of the surface pressure or as a function of the logarithm of the area. Equations 99 and 100 respectively define exactly the elasticity and exactly the compressibility (A is the area of the film and π is the surface pressure)[37, 38], but are defined in GDA by equations 101 and 102: the elasticity corresponds to the slope at one point in the graphic pressure as a function of naperian logarithm of the area while the compressibility corresponds to the inverse of the slope at one point in the graphic pressure as a function of naperian logarithm of the area).

$$\varepsilon = -A\frac{d\pi}{dA} \quad \text{Eq. 99} \qquad \text{and} \qquad C_m = -\frac{d\ln A}{d\pi} \quad \text{Eq. 100}$$

$$E = -\frac{\text{SurfacePressure}_{n-1} - \text{SurfacePressure}_{n+1}}{\text{LnArea}_{n-1} - \text{LnArea}_{n+1}} \qquad \text{Eq. 101}$$

and

$$C = \frac{1}{E} = -\frac{1}{\dfrac{\text{SurfacePressure}_{n-1} - \text{SurfacePressure}_{n+1}}{\text{LnArea}_{n-1} - \text{LnArea}_{n+1}}} \qquad \text{Eq. 102}$$

These plots are only accessible from the file menu and Figure 25 shows two typical screens after indicating the limits.

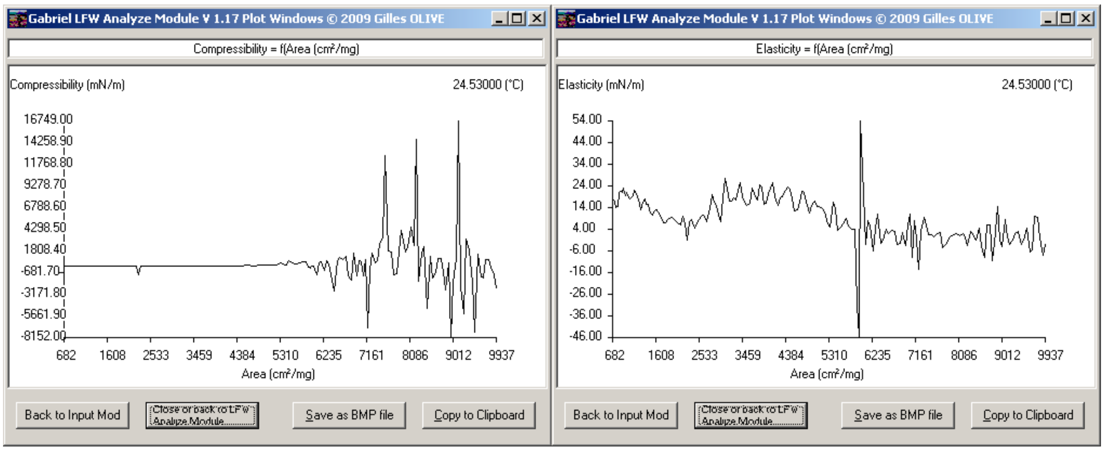
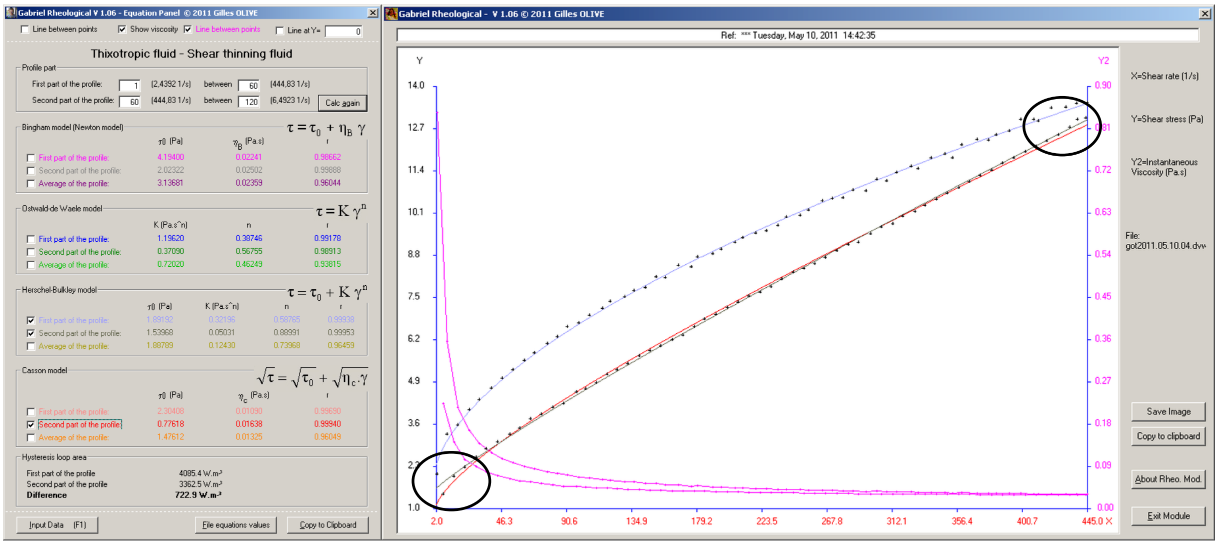


Fig. 25 : Typical plots of the compressibility and the elasticity as a function of the naperian logarithm of the area from the LFW Analyze Module.

4.10. Rheological Analyze Module (actual version: 1.06)

This module will be shortly described as almost all the algorithms have already been given. After loading a DVW file from Bohlin CVO120 and removing null or negative point (Edition, Remove negative value in Grid1 (X,Y) in the menu in the Input Module), the two screens presented in Figure 26 will appear. As the number of information given by GDA in this module is high, the same principle as the Two Variables Statistic Module is used: an equations panel and a graphical window. The last one is the same and calls no comments. The first line of the equations panel does not call any remark, like the last one as they are the same as in the Two Variables Statistic Module.



Gabriel Rheological V 1.06 - Equation Panel © 2011 Gilles OLIVE

☐ Line between points ☑ Show viscosity ☑ Line between points ☐ Line at Y= 0

Thixotropic fluid - Shear thinning fluid

Profile part

First part of the profile:	1	(2,4392 1/s)	between	60	(444,83 1/s)
Second part of the profile:	60	(444,83 1/s)	between	120	(6,4923 1/s)

Bingham model (Newton model) $\tau = \tau_0 + \eta_B\ \gamma$

	τ0 (Pa)	η_B (Pa.s)	r
☐ First part of the profile:	4.19400	0.02241	0.98662
☐ Second part of the profile:	2.02322	0.02502	0.99888
☐ Average of the profile:	3.13681	0.02359	0.96044

Ostwald-de Waele model $\tau = K\ \gamma^n$

	K (Pa.s^n)	n	r
☐ First part of the profile:	1.19620	0.38746	0.99178
☐ Second part of the profile:	0.37090	0.56755	0.98913
☐ Average of the profile:	0.72020	0.46249	0.93815

Herschel-Bulkley model $\tau = \tau_0 + K\ \gamma^n$

	τ0 (Pa)	K (Pa.s^n)	n	r
☑ First part of the profile:	1.89192	0.32196	0.58765	0.99938
☑ Second part of the profile:	1.53968	0.05031	0.88991	0.99953
☐ Average of the profile:	1.88789	0.12430	0.73968	0.96459

Casson model $\sqrt{\tau} = \sqrt{\tau_0} + \sqrt{\eta_c . \gamma}$

	τ0 (Pa)	η_c (Pa.s)	r
☐ First part of the profile:	2.30408	0.01090	0.99690
☑ Second part of the profile:	0.77618	0.01638	0.99940
☐ Average of the profile:	1.47612	0.01325	0.96049

Hysteresis loop area

First part of the profile	4085.4 W.m-³
Second part of the profile	3362.5 W.m-³
Difference	**722.9 W.m-³**

Gabriel Rheological - V 1.06 © 2011 Gilles OLIVE

Ref: *** Tuesday, May 10, 2011 14:42:35

X=Shear rate (1/s)

Y=Shear stress (Pa)

Y2=Instantaneous Viscosity (Pa.s)

File: got2011.05.10.04.dvw

Fig. 26 : Graphical window and Equations Panel of the Rheological Analyze Module.

During the transfer from the Input Module to this module, GDA searches the profile type of the shear rate: only up, only down or up and down and at the same time each part of the profile. In case of up-down profile, the thixotropic or rheoplexic behavior of the fluid is determined by examining which part of the profile, taken in the middle of each part, is above. The results are written in the second line of the first frame. In the second line the user will also find the type of the fluid (Newtonian, shear thinning or shear thickening), the determination of this being based on the behavior index n of the Herschel-Bulkley model of the whole profile. Four frames give the results for the Bingham/Newton model (linear $\tau = \tau_0 + \eta_B\ \gamma / \ \tau = \eta\ \gamma$), the Ostwald-de Waele model (power law $\tau = K\ \gamma^n$), the Herschel-Bulkley model (three parameters power law $\tau = \tau_0 + K\ \gamma^n$) and the Casson model (square root

$\sqrt{\tau} = \sqrt{\tau_0} + \sqrt{\eta_C \gamma}$). All these models have been explained in the Two Variables Statistic Module section and use the internal CLS subroutine after a variable change.

The reader will remark that every part of the rheogram can be modeled independently. On the rheogram of the kefir (Figure 26), he can also see that the second part of the profile can be modeled with two models, the Herschel-Bulkley model suitable at high shear rate while the Casson model is more appropriate for low shear rate (black circles in Figure 26).

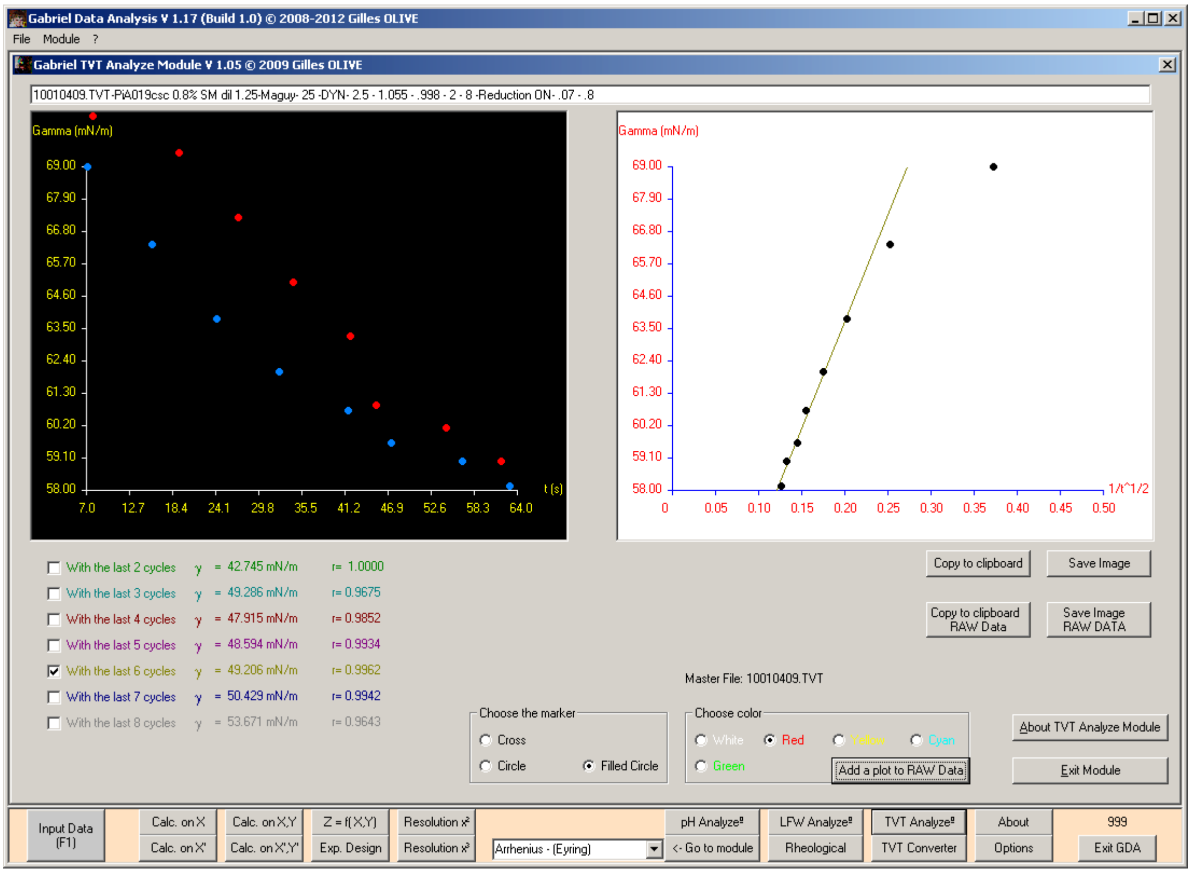
The last part is devoted to the hysteresis loop area, that uses the trapezes method (Equation 29) already described in the Two Variables Statistic Module chapter. The area is calculated for each part of the profile then the second part is subtracted of the first.

4.11. TVT Analyze Module (actual version: 1.05)

The aim of the Lauda TVT 1 apparatus and of this module is the determination of surface tension γ. Schematically the apparatus measures the fall of a drop.[37, 39] One of the parameters that can be calculated from the curve γ as a function of the time (blue circle in the left part of the Figure 27) is the equilibrium tension, that expresses the maximum surface-active power that a surface-active agent can develop. From the absorption kinetics curves the equilibrium tension γ_e (or γ_∞) can be extrapolated and can be considered as a tension for an infinite time. To estimate this tension, a variable change is needed, and the curve becomes the surface tension γ as a function of the inverse of the time square root ($1/\sqrt{t}$) that becomes a straight line. The ordinate at the origin (which corresponds to an infinite time) found by the internal CLS subroutine will give directly the equilibrium tension (right and lower part of Figure 27).

In fact, as can be seen in Figure 27, GDA will calculate the values for each number of last cycles (or point) and also gives the correlation coefficient, so the user can choose the best fit.

As a comparison, some other TVT files can be added to the RAW Data windows (red circle in the left part of the Figure 27). But the calculations are only performed on those loaded from the Input Module (Figure 27).



Gabriel Data Analysis V 1.17 (Build 1.0) © 2008-2012 Gilles OLIVE

Gabriel TVT Analyze Module V 1.05 © 2009 Gilles OLIVE

10010409.TVT-PiA019csc 0.8% SM dil 1.25-Maguy- 25 -DYN- 2.5 - 1.055 - .998 - 2 - 8 -Reduction ON- .07 - .8

□ With the last 2 cycles γ = 42.745 mN/m r= 1.0000

□ With the last 3 cycles γ = 49.286 mN/m r= 0.9675

□ With the last 4 cycles γ = 47.915 mN/m r= 0.9852

□ With the last 5 cycles γ = 48.594 mN/m r= 0.9934

☑ With the last 6 cycles γ = 49.206 mN/m r= 0.9962

□ With the last 7 cycles γ = 50.429 mN/m r= 0.9942

□ With the last 8 cycles γ = 53.671 mN/m r= 0.9643

Master File: 10010409.TVT

Choose the marker: ○ Cross ○ Circle ● Filled Circle

Choose color: ○ White ● Red ○ Yellow ○ Cyan ○ Green

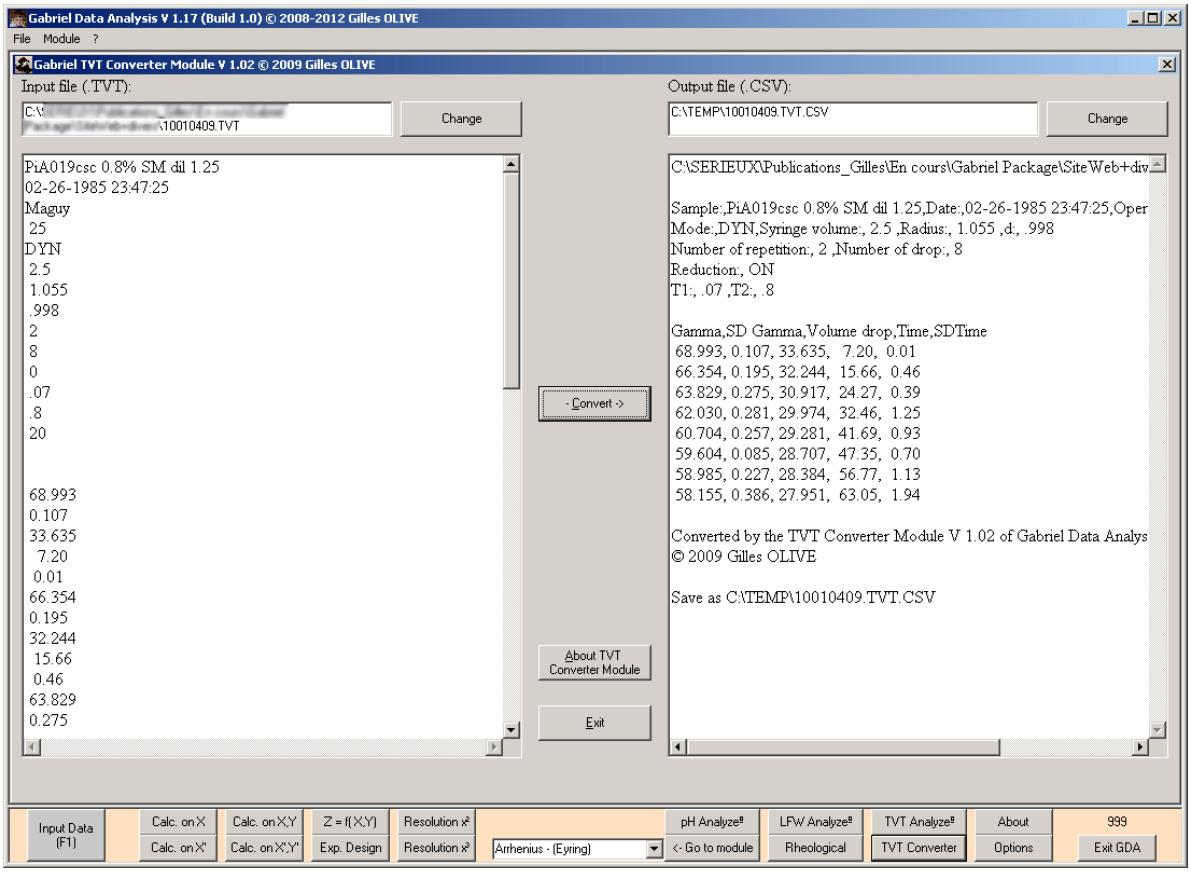
Arrhenius - (Eyring)

Fig. 27 : TVT Analyze Module screen of GDA.

4.12. TVT Converter Module (actual version: 1.02)

The conversion of a file from scientific apparatus to ASCII is very useful as these files can then be used in different software for editing, other calculations, etc ...

This module converts files from Lauda TVT 1 equipped with Lauda Drop-Volume-Tensiometer TVT-1 Version 2.0 software to a CSV file usable in Microsoft Excel®.



Input file (.TVT):

```
C:\[redacted]\10010409.TVT
```

```
PiA019csc 0.8% SM dil 1.25
02-26-1985 23:47:25
Maguy
 25
DYN
 2.5
 1.055
 .998
 2
 8
 0
 .07
 .8
 20


 68.993
 0.107
 33.635
   7.20
  0.01
 66.354
 0.195
 32.244
  15.66
  0.46
 63.829
 0.275
```

Output file (.CSV):

```
C:\TEMP\10010409.TVT.CSV
```

```
C:\SERIEUX\Publications_Gilles\En cours\Gabriel Package\SiteWeb+div

Sample:,PiA019csc 0.8% SM dil 1.25,Date:,02-26-1985 23:47:25,Oper
Mode:,DYN,Syringe volume:, 2.5 ,Radius:, 1.055 ,d:, .998
Number of repetition:, 2 ,Number of drop:, 8
Reduction:, ON
T1:, .07 ,T2:, .8

Gamma,SD Gamma,Volume drop,Time,SDTime
 68.993, 0.107, 33.635,   7.20,  0.01
 66.354, 0.195, 32.244,  15.66,  0.46
 63.829, 0.275, 30.917,  24.27,  0.39
 62.030, 0.281, 29.974,  32.46,  1.25
 60.704, 0.257, 29.281,  41.69,  0.93
 59.604, 0.085, 28.707,  47.35,  0.70
 58.985, 0.227, 28.384,  56.77,  1.13
 58.155, 0.386, 27.951,  63.05,  1.94

Converted by the TVT Converter Module V 1.02 of Gabriel Data Analys
© 2009 Gilles OLIVE

Save as C:\TEMP\10010409.TVT.CSV
```

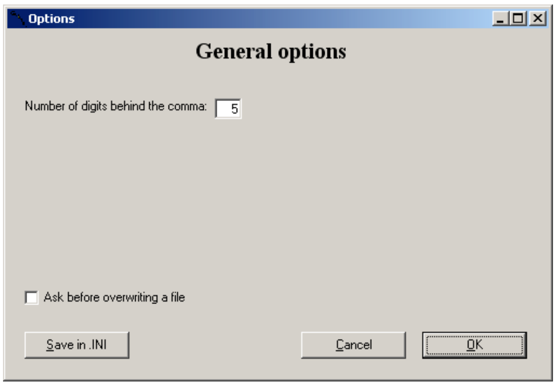
Arrhenius - (Eyring)

Fig. 28 : TVT Converter Module screen.

As can be seen in Figure 28 the left part is devoted to the initial file (at the top, the pathway while the middle part is the file itself in ASCII) and the right part to the converted file (at the top again the pathway, the middle part the file as it will be saved when the button Convert is pressed).

4.13. Options

The number of digit after the dot can be chosen for some modules (One Variable Statistic, Two Variables Statistic, LFW analyze modules, etc...) (Figure 29).



General options

Number of digits behind the comma: 5

□ Ask before overwriting a file

Fig. 29 : The options panel.

It is sometimes annoying when saving, to have a message that asks for overwriting a file, especially when entering the data. The user can check the checkbox if he wants to have a warning message in case of overwriting (Figure 29).

Those information can be saved in .INI file and be used at the next launch of GDA.

4.14. Examples

In this section, some examples will be described where GDA has been used.

In 2009, Freschi used GDA for the determination of equilibrium surface tension of the different parts of the egg yolk.[40] This can help to find new applications for those compounds, in particular in the food industry.

Fatty acid sugar esters are non-ionic, biodegradable, non toxic, non skin irritant, odourless and tasteless surfactants which allow their use in the food or the cosmetic industries.[41-43] In recent years, lipase-catalyzed synthesis of sugar esters has emerged as an interesting alternative of the chemical synthesis because of milder conditions using less toxic solvents.[44-48] The initial water content is important in the enzymatic reaction in several ways. Modeling the different parameters plays an important role for industrial processes. In this way, the author published a paper dealing with the influence of the initial water content on the conversion rate of the lauric acid, that can be modelled by a Herschel-Bulkley three-parameter power law equation, in which the first coefficient is related to the experimental conditions if no water is present and the third (the exponent) corresponds to the "efficiency" of the water (for decreasing the conversion rate).[49] In this paper, the litterature data have been fitted too. In a submitted publication dealing with the study of some parameters of the lipase-catalyzed reaction, the authors found that the influence of the reaction time can be modeled by a hyperbolic correlation.[50]

Now, some useful equations will be given to demonstrate the power of GDA. All data are taken from the CRC Handbook of chemistry and physics.[51]

The first equation concerns the pressure vapor of ice. The curve can be well fitted only by a polynomial with a degree of 6, given at equation 103 (the pressure is in Pascal while the temperature is in the range of 0 to -80 °C).

$$P = 610.9070 + 50.0988\ t + 1.8416\ t^{2} + 00387\ t^{3} + 4.8592.10^{-04}\ t^{4}$$

$$+ 3.3852.10^{-06}\ t^{5} + 1.0050.10^{-08}\ t^{6} \qquad \text{Eq. 103}$$

$$r = 1{,}00000 - \text{SEE} = 0{,}11528$$

The second equation corresponds to the refractive index of water as a function of the temperature. In this case a polynomial with an order of 4 is enough (equation 104, with the temperature ranging from 0 to 100 °C)

$$n_{D}^{t} = 1.33432 - 2.60198.10^{-06}\ t - 2.56868^{-06}\ t^{2} + 1.52040.10^{-08}\ t^{3} - 4.98252^{-11}\ t^{4} \qquad \text{Eq. 104}$$

$$r = 1{,}00000 - \text{SEE} = 1{,}33.10^{-05}$$

And the last one is a three dimensional analysis of the water electrical conductivity (σ in µS.cm^{-1}), because the conductivity depends on the temperature (in the range of 0 to 600 °C) and the pressure (in the range of 50 to 600 MPa).[51] The results of the analysis by the Z = f(X, Y) Module are given by the equations 105 to 108. As can be seen, the best correlation is obtained by equation 105, corresponding to the product of the temperature and the pressure.

$$\sigma = -2.30810 + 2.29162.10^{-4}\ t \times P \qquad r = 0.94156 \qquad \text{Eq. 105}$$

$$\sigma = -14.71085 + 0.05611\ t + 0.05269\ P \qquad s_{\sigma} = 20.43166\ ;\ s_{\sigma.tP} = 13.03144\ ;\ R_{\sigma.tP} = 0.7702 \qquad \text{Eq. 106}$$

$$\sigma = -0.90389 + 0.05955\ t \qquad r = 0.56499 \qquad \text{Eq. 107}$$

$$\sigma = -3.21074 + 0.05601\ P \qquad r = 0.55754 \qquad \text{Eq. 108}$$

5. Conclusion

In this paper the equations, the data sets and the algorithms that are in use in the different software belonging to the Gabriel package have been detailed. All this software can be used in research or educational.

Now new functionalities will be added like for example, an HLB calculator using the Davie's group number for Gabriel Calculatrice Savon or new fit quality indicators in Gabriel Data Analysis. And more particularly the change of a programming language to have a better compatibility with all computers will be undertaken.

6. Acknowledgement

The author would like to thank Paulette Desy for the helpful English correction of this manuscript. This publication is dedicated to my soon Gabriel.

7. References

[1] M. J. Frisch; G. W. Trucks; H. B. Schlegel; G. E. Scuseria; M. A. Robb; J. R. Cheeseman; J. A. Montgomery Jr.; T. Vreven; K. N. Kudin; J. C. Burant; J. M. Millam; S. S. Iyengar; J. Tomasi; V. Barone; B. Mennucci; M. Cossi; G. Scalmani; N. Rega; G. A. Petersson; H. Nakatsuji; M. Hada; M. Ehara; K. Toyota; R. Fukuda; J. Hasegawa; M. Ishida; T. Nakajima; Y. Honda; O. Kitao; H. Nakai; M. Klene; X. Li; J. E. Knox; H. P. Hratchian; J. B. Cross; V. Bakken; C. Adamo; J. Jaramillo; R. Gomperts; R. E. Stratmann; O. Yazyev; A. J. Austin; R. Cammi; C. Pomelli; J. W. Ochterski; P. Y. Ayala; K. Morokuma; G. A. Voth; P. Salvador; J. J. Dannenberg; V. G. Zakrzewski; S. Dapprich; A. D. Daniels; M. C. Strain; O. Farkas; D. K. Malick; A. D. Rabuck; K. Raghavachari; J. B. Foresman; J. V. Ortiz; Q. Cui; A. G. Baboul; S. Clifford; J. Cioslowski; B. B. Stefanov; G. Liu; A. Liashenko; P. Piskorz; I. Komaromi; R. L. Martin; D. J. Fox; T. Keith; M. A. Al-Laham; C. Y. Peng; A. Nanayakkara; M. Challacombe; P. M. W. Gill; B. Johnson; W. Chen; M. W. Wong; C Gonzalez; J. A. Pople, "Gaussian 03", V. Revision C.02, **2004** (Gaussian, Inc., Wallingford CT).

[2] Jay William Ponder, "Tinker Home Page.", http://dasher.wustl.edu/tinker/ (acceded 12th of January 2010).

[3] Alex A. Granovsky, "Firefly Home Page.", http://classic.chem.msu.su/gran/gamess/index.html (acceded 12th of January 2010).

[4] Masahiko Suenaga, "Facio Home Page.", http://www1.bbiq.jp/zzzfelis/Facio.html (acceded 12th of January 2010).

[5] Sylvia Pietri; François Le Moigne; Malvina Miollan; Marcel Culcasi, "Utilisation d'Aminophosphonates comme Marqueurs de pH en RMN du 31P", *Patent*, **1999**, *WO9947527*.

[6] Sylvia Pietri; Malvina Miollan; Sophie Martel; François Le Moigne; Bruno Blaive; Marcel Culcasi, "α- and β-Phosphorylated Amines and Pyrrolidines, a New Class of Low Toxic Highly Sensitive ^{31}P NMR pH Indicators. Modeling of pKa and chemical shift values as a function of substituents", *Journal of Biological Chemistry*, **2000**, *275 (26)*, 19505-19512.

[7] Sylvia Pietri; Sophie Martel; Marcel Culcasi; Marie-Christine Delmas-Beauvieux; Paul Canioni; Jean-Louis Gallis, "Use of Diethyl (2-methylpyrrolidin-2-yl) phosphonate as a Highly Sensitive Extra-and Intracellular ^{31}P NMR pH Indicator in Isolated Organs", *Journal of Biological Chemistry*, **2001**, *276 (3)*, 1750-1758.

[8] Sophie Martel, "Les aminophosphonates: des marqueurs de pH en RMN ^{31}P. De l'étude physico-chimique à la biologie.," Chimie, Université de Provence (Aix-Marseille I), 2002.

[9] Sophie Martel; Jean-Louis Clément; Agnès Muller; Marcel Culcasi; Sylvia Pietri, "Synthesis and ^{31}P NMR Characterization of New Low Toxic Highly Sensitive pH Probes Designed for In Vivo Acidic pH Studies", *Bioorganic and Medicinal Chemistry*, **2002**, *10* 1451-1458.

[10] Gaëlle Gosset; Sophie Martel; Jean-Louis Clement; Bruno Blaive; Gilles Olive; Marcel Culcasi; Roselyne Rosas; André Thevand; Sylvia Peitri, "Nouveaux marqueurs de pH utilisables en RMN du ^{31}P : étude de la relaxation longitudinale en fonction de la structure chimique, de la température, du pH et du milieu biologique (New highly sensitive pH probes for ^{31}P NMR spectroscopy: study of spin-lattice relaxation time as a function of molecular structure, temperature, pH and biological medium)", *Comptes Rendus Chimie*, **2008**, *11 (4-5)*, 541-552.

[11] Gaelle Gosset; Michel Satre; Bruno Blaive; Jean-Louis Clement; Jean-Baptiste Martin; Marcel Culcasi; Sylvia Pietri, "Investigation of subcellular acidic

compartments using alpha-aminophosphonate P-31 nuclear magnetic resonance probes", *Analytical Biochemistry*, **2008**, *380 (2)*, 184-194.

[12] Bruno Blaive, "COURBE: Numerical analysis of points and functions - Resolution, regression, integration", **1992-2007** (University of Aix-Marseille III, Marseilles).

[13] Sujoy Mukherjee; Cancan Huang; Francisco Guerra; Ke Wang; Eric Oldfield, "Thermodynamics of Bisphosphonates Binding to Human Bone: A Two-Site Model", *Journal of the American Chemical Society*, **2009**, *131 (24)*, 8374-8375.

[14] O. O. Fasina; H. Hallman; M. Craig-Schmidt; C. Clements, "Predicting Temperature-Dependence Viscosity of Vegetable Oils from Fatty Acid Composition", *Journal of the American Oil Chemical Society*, **2006**, *83 (10)*, 899-903.

[15] Juyoung Kim; Deok Nyun Kim; Sung Lee Lee; Sang-Ho Yoo; Suyong Lee, "Correlation of fatty acid composition of vegetable oils with rheological behaviour and oil uptake", *Food Chemistry*, **2010**, *118* 398-402.

[16] R. Von Bergen; E. Rogel, "A lattice fluid approach to the concept of hydrophile lipophile balance: alkylethoxylated surfactants", *Fluid Phase Equilibria*, **1998**, *153 (1)*, 63-72.

[17] Z.E. Proverbio; S.M. Bardavid; E.L. Arancibia; P.C. Schulz, "Hydrophile-lipophile balance and solubility parameter of cationic surfactants", *Colloids and Surfaces a-Physicochemical and Engineering Aspects*, **2003**, *214 (1-3)*, 167-171.

[18] Xiaowen Guo; Zongming Rong; Xugen Ying, "Calculation of hydrophile-lipophile balance for polyethoxylated surfactants by group contribution method", *Journal of Colloid and Interface Science*, **2006**, *298 (1)*, 441-450.

[19] R. C. Pasquali; M. P. Taurozzi; C. Bregni, "Some considerations about the hydrophilic-lipophilic balance system", *International Journal of Pharmaceutics*, **2008**, *356 (1-2)*, 44-51.

[20] Valeria Verdinelli; Paula V. Messina; Pablo C. Schulz; Bruno Vuano, "Hydrophile-lipophile balance (HLB) of n-alkane phosphonic acids and theirs salts", *Colloids and Surfaces a-Physicochemical and Engineering Aspects*, **2008**, *316 (1-3)*, 131-135.

[21] Thomaelle, "Les lys blancs de Venus", http://lysblancsdevenus.canalblog.com/archives/les_techniques_de_base/index.html (acceded 30th of December 2010).

[22] H. Abramovic; C. Klofutar, "The temperature dependence of dynamic viscosity for some vegetable oils", *Acta Chimica Slovenia*, **1998**, *45 (1)*, 69-77.

[23] J. P. Nougier, Méthodes de calcul numérique (Paris: Masson, 1983) 19-114, 144-171 (317 pp).

[24] Murray R. Spiegel, Théorie et applications de la statistique - 875 exercices résolus., Serie Schaum (New York: McGraw-Hill, 1974) 45-88, 188-200, 217-282, 344 (358 pp).

[25] Vincent Cavaseno, Calculator programs for chemical engineers, Chemical Engineering Magazine (New York: McGraw-Hill Publications Co., 1982) 15-38, 56-62, 75-78 (328 pp).

[26] J. Vincente de Julian-Ortiz; Lionelle Pogliani; Emili Besalu, "Two-Variable Linear Regression: Modeling with Orthogonal Least-Squares Analysis", *Journal of Chemical Education*, **2010**, *87 (9)*, 994-995.

[27] Edgar D. Smith; Daniel M. Mathews, "Least Square Regression Lines. Calculations assuming a constant percent error", *Journal of Chemical Education*, **1967**, *44 (12)*, 757-759.

[28] Leslie Glasser, "Dealing with Outliers: Robust, Resistant Regression", *Journal of Chemical Education*, **2007**, *84 (3)*, 533-534.

[29] Glen Mullineux, "Non-linear least squares fitting of coefficients in the Herschel-Bulkley model", *Applied Mathematical Modeling*, **2008**, *32 (12)*, 2538-2551.

[30] Martial Chabanel, Thermodynamique chimique (Paris: Ellipses, 1986) 244-277 (284 pp).

[31] "Wikipedia: The complex number", http://en.wikipedia.org/wiki/Complex_number (acceded 12th of January 2011).

[32] Marc Gourion; Jacques Chevalet; Christian Lixi, Mathématiques. Terminale D, Collection Marc Gourion (Paris: Fernand Nathan, 1983) 149-176 (286 pp).

[33] "Wikipedia: The complex plane", http://en.wikipedia.org/wiki/Complex_plane (acceded 12th of January 2011).

[34] "Wikipedia: La méthode de Sotta", http://fr.wikipedia.org/wiki/Méthode_de_Sotta (acceded 12th of January 2011).

[35] Hasan Togrul; Nurhan Arslan, "Mathematical model for prediction of apparent viscosity of molasses", *Journal of Food Engineering*, **2004**, *62* 281-289.

[36] "Wikipedia: Graphe de Gran", http://fr.wikipedia.org/wiki/Graphe_de_Gran (acceded 18th of October 2011).

[37] Nadia Innocente; Cesare Corradini; Christophe Blecker; Michel Paquot, "Dynamic surface properties of the Proteose-Peptone fraction of bovine milk", *Journal of Dairy Science*, **1998**, *81 (7)*, 1833-1839.

[38] Abdourahamane Balla, "Etude des propriétés interfaciales du gluten et des protéines du Sorgho en vue de la panification," Thèse de doctorat, Communauté française de Belgique, Faculté des Sciences Agronomiques de Gembloux, 1999.

[39] Gaoussou Karamoko, "Contribution à l'étude des propriétés physico-chimiques de la fraction proteose-peptone," Travail de fin d'études, Année académique **2006-2007**, Communauté française de Belgique, Faculté des Sciences Agronomiques de Gembloux, 2007.

[40] Jérôme Freschi, "Etude des propriétés physicochimiques de différentes fractions du jaune d'oeuf," Travail de fin d'études, Année académique **2008-2009**, Communauté française de Belgique, Faculté des Sciences Agronomiques de Gembloux, 2009.

[41] S. Soultani; J.-M. Engasser; M. Ghoul, "Effect of acyl donor chain length and sugar/acyl donor molar ratio on enzymatic synthesis of fatty acid fructose esters", *Journal of Molecular Catalysis B: Enzymatic*, **2001**, *11* 725-731.

[42] Gülten Sekeroglu; Sibel Fadiloglu; Esra Ibanoglu, "Production and characterisation of enzymatically produced lauric acid esters of fructose", *Journal of the Science of Food and Agriculture*, **2002**, *82 (13)*, 1516-1522.

[43] S. Sabeder; M. Habulin; Z. Knez, "Lipase-catalyzed synthesis of fatty acid fructose esters", *Journal of Food Engineering*, **2006**, *77* 880-886.

[44] D. Coulon; A. Ismail; M. Girardin; B. Rovel; M. Ghoul, "Effect of different biochemical parameters on the enzymatic synthesis of fructose oleate", *Journal of Biotechnology*, **1996**, *51* 115-121.

[45] D. Coulon; M. Girardin; M. Ghoul, "Enzymic synthesis of fructose monooleate in a reduced pressure pilot scale reactor using various acyl donors", *Process Biochemistry*, **1999**, *34* 913-918.

[46] F. Chamouleau; D. Coulon; M. Girardin; M. Ghoul, "Influence of water activity and water content on sugar esters lipase-catalyzed synthesis in organic media", *Journal of Molecular Catalysis B: Enzymatic*, **2001**, *11* 949-954.

[47] Salvator Piccicuto; Christophe Blecker; Jean-Christophe Brohée; Georges Lognay; Claude Deroanne; Michel Paquot; Michel Marlier, "Les esters de sucres : voies de synthèse et potentialités d'utilisation", *Biotechnologie Agronomie Société et Environnement*, **2001**, *5 (4)*, 209-219.

[48] Benoît Moreau; Georges C. Lognay; Christophe Blecker; Jean-Christophe Brohée; Florence Chéry; Patrick Rollin; Michel Paquot; Michel Marlier, "Synthesis of novel D-glucuronic acid fatty esters using *Candida antartica* lipase in *tert*-butanol", *Biotechnology Letters*, **2004**, *26* 419-424.

[49] Gilles Olive; Gabriela A. Pompeu Torezan; Christophe Blecker, "Influence of the water content in the synthesis of fructose laurate by immobilized lipase B of *Candida antarctica* (Novozym 435)", *Bulletin de la Société Royale des Sciences de Liège*, **2012**, *81* 82-89.

[50] Gilles Olive; Gabriela A. Pompeu Torezan; Christophe Blecker, "Synthèse enzymatique d'esters de fructose (Enzymatic synthesis of fructose esters)", *Comptes Rendus Chimie*, **2012**, *15 (11-12)*, 1037-1047.

[51] CRC Handbook of chemistry and physics, ed. David R. Lide, 88th Edition ed. (Boca Raton: CRC Press, Taylor & Francis Group, 2007-2008) 5-71, 76-10, 10-251.