Annals of West University of Timisoara,

Series Chemistry **12 (3)** (2003) 285-292

MODELING OF STATISTICAL KINETICS FAST REDOX REACTIONS DATA

Mihaela Ligia UNGUREŞAN[a*], Lorentz JÄNTSCHI[a]

[a] Technical University of Cluj-Napoca, Faculty of Science and Engineering of Materials, Chemistry Department, Muncii Street, 103-105, Cluj, 3400, ROMANIA

SUMMARY

The present paper is focus on modeling of statistical data processing with applications in field of material science and engineering. A new method of data processing is present and applies on a set of kinetic data for the inner sphere redox reaction between copper (II) and thiosulfate ions. The method allows to identify the correlations between data sets and to exploit them later in statistical study kinetic rapid redox reactions. An algorithm for computing data was implements in preprocessed hypertext language (PHP), an interface in hypertext markup language for them was also realizes and put onto academicdirect.ro educational web server, and it is accessible via http protocol at the address http://academicdirect.ro/~lori/research/regression/linear/v1.5/.

Keywords: Modeling, fast redox reactions, automat processing of data

INTRODUCTION

In field of statistical data processing it exist a large set of software to compute and fit the regressions, but few are free. Even for free software, another problem it appear, operating system license and portability of the software. As example, to use well-known Microsoft Excel software, you must have Microsoft Windows and Microsoft Excel license and portability of Excel program is restricted to Windows platform. To import Excel files in another programs or platforms, conversion modules are need, and conversion is not totally in most of the cases. Platform independent and free software is a real alternative to this. First step to build totally free software is to install a free operating system. Unix-like operating systems are known to free, but even here exist licensed software. In order to select a totally free Unix-like operating system, best offer come from BSD family. The most secure and license check for installed software is NetBSD. The NetBSD detect so called "license agreement" and do not permit to install a software if the software contain

unacceptable license agreement (different form free) and software can be installed only if the acceptance is explicitly stipulated by the user in configuration files of the system. Another advantage of NetBSD operating system is his huge portability under various hardware platforms from i386 family to Sun and Macintosh machines. By another hand, most full featured operating system of BSD family is FreeBSD. One of the advantages of the FreeBSD operating system is his software portability. With adequate packages, under FreeBSD, can be executed DOS, Windows, Linux and Sun-OS programs. Another advantage of FreeBSD system comes from easiest to install and use it. Once you have an operating system installed, the next step is to choose a proper programming language for software developing. Here, some major questions must be answer. In terms of programming, portability of resulted program can be a problem. As example, if we are chose to implement the algorithm in Visual Basic, the execution of the program is restricted to Windows machines. If Perl is our choice, a Unix-based machine is necessary to run program. The advantage of PHP programs consist in his portability under most of operating system platforms and internal compilation feature that do not necessity the compilation "by hand" from the user. The disadvantage can be same internal compilation that consumes supplementary time in execution. However, this disadvantage can be partially eliminates through installing a PHP proxy, which store compiled programs and next execution of the unmodified program use this compiled binary. In terms of program developing PHP is easy to use, the language borrow syntax from C, Pascal, Basic and Perl, but do not borrow the complex declaration syntax from them. The pointer mechanism is absolute. Thus, a variable used as a string, can be exploited as an integer or real if the value represent a number such that. Class constructing is also available and PHP posses a strong library of database connectivity. A very easy mechanism to link PHP scripts to HTML scripts make PHP language to be one of the best. Shell executing commands make PHP a useful platform for system administration (PhpSysInfo, WebAdmin, PhpMyAdmin, PhpPgAdmin). PHP programs are putting on a web's server data folder and executed by them using PHP module. The output of the PHP program is in HTML style and can be viewed by any web client (Microsoft Internet Explorer, Mozzilla, Opera, Netscape, Konkueror).

MATERIALS AND METHODS

Many statistical procedures for processing data are available [1]. Most of them offer a voluble set of possibilities and variants, but which one to consider them? That is not an easy question and the frequent answer is: that is choice of analyst [2,3]. Data mining technology offer in this area of knowledge some answers, but not a complete answer [4]. By other hand, to interpret experiment results, data need to be well processed [5]. Structure investigations are frequently combining with statistical processing [6]. In most of cases, best results are obtained with specific procedures in contrast to general numeric algorithms [7,8]. Modeling of structure is benefit to property predictions [9,10]. Nonstandard statistical

evaluation procedures then are helpful [11]. The design of statistical processing program is depicting in Figure 1:

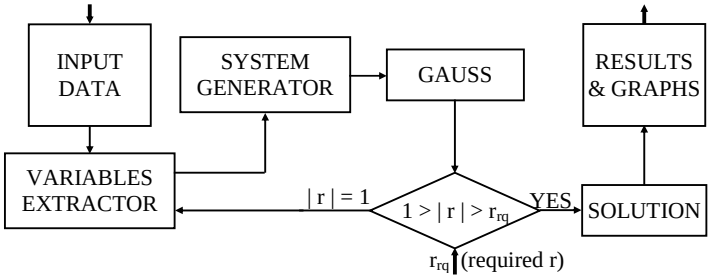


Figure 1. Program design scheme

The INPUT module read a text format data, process input data, split it into rows and columns. The rows represent the values of variables and columns represent the variables.

The INPUT module passes data in matrix form to the VARIABLES EXTRACTOR module. This module starts with pairs of columns from data matrix; continue with sets of three columns and so on, until entire set is browse. For every set, pass it to the SYSTEM GENERATOR. Another input in VARIABLES EXTRACTOR module comes from DECISION module. If the DECISION module detects a correlation of absolute coefficient 1 then it passes to the VARIABLES EXTRACTOR the first variable label from the regression equation and for the next extractions of variables, the passed variable is marked as dependent and is not passes never again to the SYSTEM GENERATOR module. The SYSTEM GENERATOR computes average means, needed by GAUSS module. If answer of algorithm solving is *undetermined system* and null variable is x_{n_cols} then GAUSS module solve determined system of *n_cols* order. When system is solves, a unique solution is found. Then, system extended matrix contain at column n_cols the coefficients of regression equation:

$$a_1 \cdot x_1 + ... + a_i \cdot x_i + ... + a_{n\_cols} \cdot x_{n\_cols} + a_{n\_cols+1} = 0 \,;\; a_{n\_cols+1} = -1 \qquad (1)$$

where the coefficients a_{n_cols+1} and a_{n_cols+1} are resulted regression coefficients. Note that equation (1) is in implicit form; to obtain an explicit form is necessary to extract dependent variable from (1). The last coefficient is assign to -1. At the end of module SOLUTION it result an implicit linear regression equation between given variables through his values in columns (eq. 1). Equation 1 could be exploiting to obtain explicit linear regression equations for each variable, which has no null coefficient a_i:

$$\hat{x}_i = \left(\frac{a_1}{-a_i}\right) \cdot x_1 + ... + \left(\frac{a_{i-1}}{-a_i}\right) \cdot x_{i-1} + \left(\frac{a_{i+1}}{-a_i}\right) \cdot x_{i+1} + ... + \left(\frac{a_{n\_cols}}{-a_i}\right) \cdot x_{n\_cols} + \left(\frac{a_{n\_cols+1}}{-a_i}\right) \qquad (2)$$

Sum of residues S_i, and with data replacements, Q_i can be now evaluates:

$$S_i = \left(\frac{a_{n\_cols+1}}{a_i} + \sum_{j=1}^{n\_cols} \frac{a_j}{a_i} \cdot x_i \right)^2 ; Q_i = \sum_{k=1}^{n\_rows} \left(\frac{a_{n\_cols+1}}{a_i} + \sum_{j=1}^{n\_cols} \frac{a_j}{a_i} \cdot data[k,i] \right)^2 / M[i,i] \quad (3)$$

To make independence of measurement unit and measure order, values S_i are divided with own sum of squares of variable measurements (M [i,i] from INPUT module, equation 1). Final equation Q_i, are with substitution x_i = data [k,i], $1 \leq k \leq n\_rows$. Q_i express relative residues of variable x_i when variable x_i is assumed to be dependent of independent variables x_1, …, x_{i-1}, x_{i+1}, x_{cols}. Note that the dependence and independence statistical concept is hard to prove in practical situations, but will see later, can be decelerated. For a good correlation, Q_i should be smallest possible value. The value of Q_i is computes in SOLUTION module for every equation. Another quantitative measure for a good correlation is correlation coefficient between measured data x_i and estimated values $\hat{x}_i$ from equation (1). The absolute value of r must be high for a good correlation. The value of r is computes in DECISION module for implicit equation and in SOLUTION module for every explicit equation. More, additionally tests are also available in other programs such as Microsoft Excel or Statsoft Statistica.

RESULTS AND DISCUSSION

A graphical interface was builds in PHP with a TEXTAREA for input data and an INPUT SUBMIT button for submitting data to the server. The server is a Free BSD Unix based server (5.0 DP1 software version) with an Apache web server (1.3.26 software version) running on. The server is hosts in educational network of Technical University of Cluj-Napoca with address 193.226.7.211 and name academicdirect.ro. The PHP language was compiles with GDI (graphical device interface) and MySQL (database server) support and the PHP software version is 4.2.3. The MySQL database server is also installed and running on and his software version is 3.23.52. The input interface is presents in Figure 2:

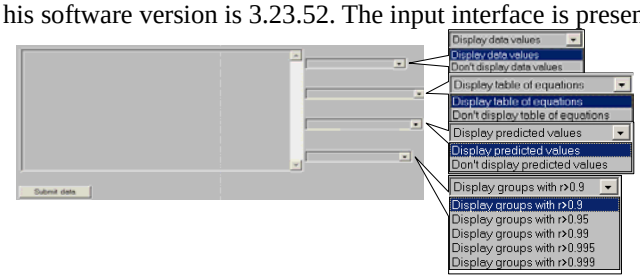


Figure 2. The output of http://193.226.7.211/~lori/research/regression/linear/v1.5/index.php

The output of the RESULTS & GRAPHS module is passes depending on user equation selection to another PHP program that makes graphical representation of regression [12]:

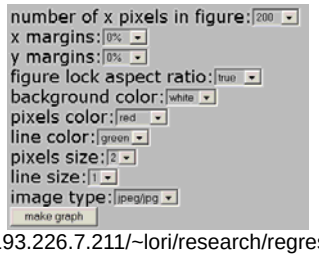


Figure 3. The output of http://193.226.7.211/~lori/research/regression/graphics/lin_plot/do.php

We have been reading the transmitted intensities I_t, at different time values, the incident intensity I_u, and according to the Lambert –Beer Law, we have calculated the extinctions E at different wave lengths and concentrations of the $Cu(ClO_4)_2$, $Na_2S_2O_3$ and $HClO_4$. It was effectuated experimental determination for concentrations of $[Cu(ClO_4)_2]$ = [0.0017, 0.05] M^{-1}, $[Na_2S_2O_3]$ = [0.005, 0.2]M^{-1} , $[HClO_4]$ = [0.008, 0.1] M^{-1}, and for length way: λ = [380, 800]nm [13]. The obtained curves for the coordinates transmitted light as a function of time (I_t = f(t)) have been translated at t = 0 and it has been calculated the solution extinction during the mixing period $E_0 = \log(I_u/I_t)_0$. The initial concentrations of the reactants are identical (or close) in all determinations, so that the influence of the pre-equilibrium on the apparent order could be manifested on its entire dimension. The graphics readings point by point are makes by digitization of the scanned figures. As a result, we have obtained an Excel file. The input data can be copy and paste into TEXTAREA from Figure 2. By pressing, submit button the input data are processed. By chousing to display data values, the input data are also displays. The properties of reaction described in Table 1. In Figure 4 are plotted the dependences from data set. For fitting, the suggested selections from PHP program are used. Figure 4 shows that no exist significance dependence between the kinetic experimental data. It observes that the experimental data was grouped in 5 horizontal linear representations:

TABLE I. Processed Data

Col.	Property	Measurement unit
1	time	ms
2	corrected time	ms
3	Extinction	no unit
4	$[Cu(ClO_4)_2]$	mol/l
5	$[HClO_4]$	mol/l
6	Temperature	K

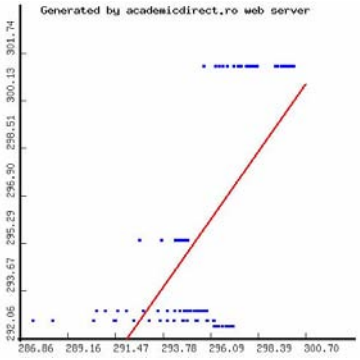


Figure 4. Regression of all data

Program computes and output the regression equations. That is a large set! If we are interested to study dependence between two variables form set, then we select the proper table from output of the program. If we are looking for very dependent variables, the program finds it and eliminates one of them from set. For all kinetic experimental data, the equation is:

$$+x_0*1.98*10^{-2}+x_1*2.56*10^{1}-x_2*3.95*10^{1}-x_3*3.81*10^{2}+x_4*1.00=+2.91*10^{2}$$

The coefficient of correlation r = 0.67813 and standard deviation s = 0.10188. In Table 2 are presents the input data, such as are showed by PHP program. For a group of 19 experimental data, [Cu(II)] = 0.05 mol/l, $[S_2O_3^{2-}]$ = 0.01 mol/l, T = 293 K (Table 2), the regression equation determined is:

$$+x_2*1.00+x_0*1.20*10^{2}=+3.85*10^{1}; r=0.90932; s=0.61886$$

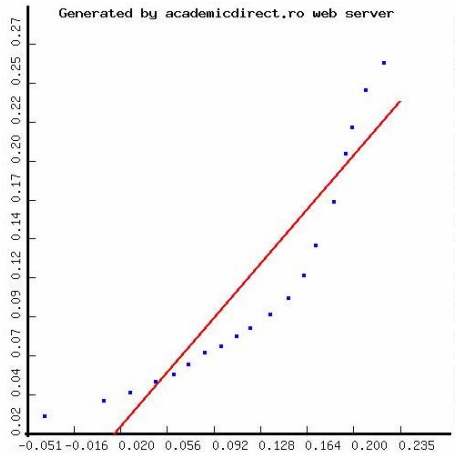


Figure 5. Extinction E = E(t) ($+x_2*1.00+x_0*1.20*10^{2}=+3.85*10^{1}$; r=0.90932; s=0.61886; n=19)

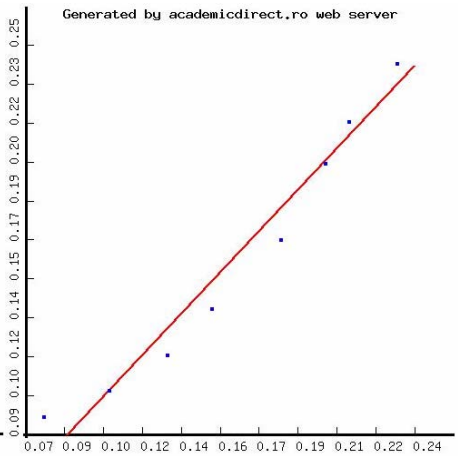
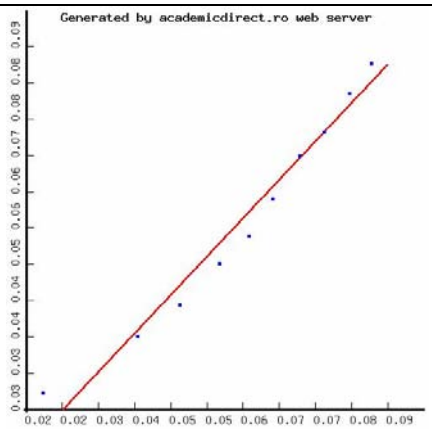


Figure 6. Extinction E = E(t) for first eight points e.g. initiation of reaction with induction moment ($+x_0*1.77*10^{-2}+x_2*1.00=+0.40$; r=0.97964; s=0.17326; n=8)

TABLE II. Input data values
(output by PHP program, [14])

1	2	3		24.136	20.14	0.0862
11.637	7.64	0.2612		25.429	21.43	0.081
13.361	9.36	0.2431		26.895	22.89	0.0741
14.568	10.57	0.2187		28.36	24.36	0.0699
15.171	11.17	0.2011		29.912	25.91	0.0623
16.292	12.29	0.1692		31.291	27.29	0.0557
18.016	14.02	0.1404		33.015	29.01	0.0508
19.137	15.14	0.1209		35.342	31.34	0.0436
20.602	16.6	0.106		37.824	33.82	0.0381
22.24	18.24	0.0951		43.359	39.36	0.028



CONCLUSION

Considering the advantages of implemented software technology (machine and operating system portability, graphical interface and database connectivity features, easiest of programs developing, free type license agreement, http capability) the programming language and the program itself is the one of the best choice now available.

Considering the figures 6 and 7, the program makes possible the study of reaction in both moments (initiating and propagation) and obtained equation prove the dependencies between time and extinction. The approximations (both linear) are good (see correlation coefficients and standard deviation) and prove that for the reaction the initiating and propagation moments can be approximates by linear dependencies:

$$Cu^{2+} + S_2O_3^{2-} \rightarrow Cu^{+} + \tfrac{1}{2}S_4O_6^{2-} \quad (4)$$

We will investigate a very large set of kinetic experimental data for given fast redox reaction using this conclusion of bi-linear behavior of (4). The initiating moment can be associated with pre-equilibrium reaction and propagation can be associated with fast decomposition of colored complex CuS_2O_3.

REFERENCES

1. Naşcu H., Jäntschi L., Hodisan T., Cimpoiu C., Câmpan G., "*Some Applications of Statistics in Analytical Chemistry*", Reviews in Analytical Chemistry, Freud Publishing House, XVIII(6), 1999, p. 409-456.
2. Jäntschi L., Unguresan Mihaela, "*MathLab. Statistical Applications*", International Conference on Quality Control, Automation and Robotics, May 23-25 2002, Cluj-Napoca, Romania, work published in volume "AQTR Theta 13", p. 194-199.
3. Jäntschi L., Nascu H., "*Numerical Description of Titration (MathCad application)*", International Conference on Quality Control, Automation and Robotics, May 23-25 2002, Cluj-Napoca, Romania, work published in volume "AQTR Theta 13", p. 259-262.
4. Jäntschi L., Unguresan Mihaela, "*Parallel processing of data. C++ Applications*", Oradea University Annals, Mathematics Fascicle, 2001, VIII, p. 105-112.
5. Marta L., Deac I. G., Fruth I., Zaharescu M., Jäntschi L., "*Superconducting Materials: Comparision Between Coprecipitation and Solid State Phase Preparation*", International Conference on Science of Materials, volume 2, August 20-22, 2000, Cluj-Napoca, Romania, p. 627-632.
6. Marta L., Deac I. G., Fruth I., Zaharescu M., Jäntschi L., "*Superconducting Materials at High Temperature: Thermal Treatment and Pb Addtition*", International Conference on Science of Materials, volume 2, August 20-22, 2000, Cluj-Napoca, Romania, p. 633-636.
7. Cimpoiu C., Jäntschi L., Hodisan T., "*A New Method for Mobile Phase Optimization in High-Performance Thin-Layer Chromatography (HPTLC)*", Journal of Planar Chromatography - Modern TLC, Springer, 11(May/June), 1998, p. 191-194.
8. Cimpoiu C., Jäntschi L., Hodisan T., "*A New Mathematical Model for the Optimization of the Mobile Phase Composition in HPTLC and the Comparision with Other Models*", Journal of Liquid Chromatography and Related Technologies, Dekker, 22(10), 1999, p. 1429-1441.
9. Diudea M., Jäntschi L., Graovac A., "*Cluj Fragmental Indices*", Math/Chem/Comp'98 (The Thirteenth Dubrovnik International Course & Conference on the Interfaces among Mathematics, Chemistry and Computer Sciences), June 22-27, 1998.
10. Jäntschi L., Katona G., Diudea M., "*Modeling Molecular Properties by Cluj Indices*", Commun. Math. Comput. Chem. (MATCH), 2000, 41, p. 151-188.
11. Sârbu C., Jäntschi L., "*Statistic Validation and Evaluation of Analytical Methods by Comparative Studies. I. Validation of Analytical Methods using Regression Analysis (in Romanian)*", Revista de Chimie, Bucuresti, 49(1), 1998, 19-24, ISSN 0034-7752.
12. Jäntschi L., Naşcu H. I., "*Free Software Development. 3. Numerical Description of Soft Acid with Soft Base Titration*", Leonardo Journal of Sciences, Vol. 1, Cluj-Napoca, 2002, p. 53-68.
13. Niac G., "*The prelucration of the obtained data from kinetic determination of isotherm's desorption*"", Scientific Bulletin of Technical University of Cluj-Napoca, Series: Applied Mathematics, Mechanics, 34, 1991, p.129 – 135.
14. Jäntschi L., "*Automat Server Side Processing of Statistical Data*", UNITECH'02 International Conference, 21-22 November 2002, Gabrovo, Bulgaria, p. 185-189.