Journal of Thermal Analysis and Calorimetry, Vol. 88 (2007) 3, 751–755

COMPENSATION LAW AGAIN

J. Norwisz and T. Musielak*

AGH – University of Science and Technology, Department of Non-Ferrous Metallurgy, Cracow, Poland

Some preliminary considerations suggest that the so-called 'compensation law' is a result of the misinterpretation of evaluation procedure. The both parameters: pre-exponential factor and activation energy are calculated from the same set of experimental data. A confidence ellipse could describe the precision of these parameters. In the case of using the least square method for straight-line parameters evaluation, we could calculate *a* axes of the ellipse, using experimental data. If the observed relationship between pre-exponential factor and activation energy agree with the 'pre-calculated' direction of main axis of confidence ellipse, we have a strong support to believe that the observed 'compensation' effect is only an artificial effect of misinterpretation. Some calculations performed for a published experimental data have confirmed these suspicions. This also, indirectly indicates that precision of such experiments is probably lower than expected.

Keywords: *apparent linear relationship, compensation effect, compensation law, confidence ellipse, covariance matrix*

Introduction

One of the frequently appearing topics in the literature devoted to high temperature kinetics is the 'compensation' law. The interesting thing is that this simple relationship between pre-exponential factor and energy activation of the Arrhenius equation is so often observed for kinetic experiments [e.g. 1–7], and rather uncommon in other fields of physics or chemistry. Some preliminary considerations suggest that the law is a result of the misinterpretation of evaluation procedure [8–11].

The long-term discussion concerning existence or (non-existence) the enigmatic law of 'linear compensation' [e.g. 12–14] omits the essential fact that the subject of assessment here is a two-dimensional random vector of linear regression equation coefficients. The subject of the paper is to indicate that the effect of 'linear compensation' is of apparent character and results from a simplified interpretation of estimation procedure for linear regression equation coefficients. It results directly from the high inaccuracy of measurement results, being even higher than the values admitted by experimenters.

Two-dimensional random variable

Density of normal k-dimensional random variable

$$z^{T}=[z_1, z_2, ..., z_k] \tag{1}$$

defines the following relationship:

$$f(z_1, z_2, ..., z_k)= \frac{1}{(2\pi)^{k/2}|M|^{1/2}}\exp\left[-\frac{1}{2}(z-\mu)^{T}M^{-1}(z-\mu)\right] \tag{2}$$

where μ is an expected value for distribution vector, M – covariance matrix, index T stands for transposition; i.e. changing matrix verses into columns. The surface area of constant distribution is a k-dimensional ellipsoid determined by constant value $C(\alpha)$

$$\left[-\frac{1}{2}(z-\mu)^{T}M^{-1}(z-\mu)\right]=\text{const.}=C(\alpha) \tag{3}$$

There is a definite probability α, that the result of estimating the random variable z will be inside the ellipsoid calculated for constant $C(\alpha)$. Because the vector z is a normal vector, the value $C(\alpha)$ has a distribution of chi-square with k degrees of freedom. It allows us to calculate the multi-dimensional probability without using tables calculated for the particular examined normal distribution. The ellipsoids determined for different values of α are similar, having the same center defined by the location of expected value point μ, and the same axis.

In the simplest case of two-dimensional random variable with expected value vector

$$\mu^{T}=[\mu_1,\mu_2] \tag{4}$$

and covariance matrix

$$M=\begin{bmatrix}\sigma_1^2 & \text{cov}\\ \text{cov} & \sigma_2^2\end{bmatrix} \tag{5}$$

* Author for correspondence: norwisz@agh.edu.pl,

the curve of density uniformity is an ellipse, whose shape and position is defined by the expected value vector μ and symmetrical covariance matrix M.

Linear regression equation coefficients as two-dimensional normal random vector

A routine procedure for processing experimental results elaboration is assessing the values of coefficients a and b of linear regression equation [e.g. 16].

$$y=ax+b \tag{6}$$

for a particular set of n pairs of experimental results (y_i, x_i), using the least squares method. It leads to obtain well-known terms allowing us to calculate the values of regression equation coefficients.

$$a=\frac{\sum_{i=1}^{n} y_i x_i-\frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n}}{\sum_{i=1}^{n} x_i^2-\frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \tag{7a}$$

$$b=\frac{\sum_{i=1}^{n} y_i}{n}-a\frac{\sum_{i=1}^{n} x_i}{n} \tag{7b}$$

It is assumed that only the results of dependent variable y_i are subject to measuring error, and the accuracy for all measurements is identical. The value of single standard measurement deviation s_0 can be evaluated on the base of residual sum for measurement results

$$s_0^2=\frac{\sum_{i=1}^{n}(y_i-ax_i-b)^2}{n-2} \tag{8}$$

as the value s_0^2 has a chi-square distribution with (n–2) degrees of freedom.

From the relationships (7a) and (7b) we can see that both regression equation coefficients are linear functions of measurement results. If experimental values are normal variables, it directly implies that the vector $[a, b]^T$ is also a normal vector. The covariance matrix for this vector can be determined according to the rule of linear error propagation, from the following equation:

$$M=\begin{bmatrix} \frac{\partial a}{\partial y_1} & \cdots & \frac{\partial a}{\partial y_n} \\ \frac{\partial b}{\partial y_1} & \cdots & \frac{\partial b}{\partial y_n} \end{bmatrix} \begin{bmatrix} s_0^2 & & & \\ & s_0^2 & & \\ & & \ddots & \\ & & & s_0^2 \end{bmatrix} \begin{bmatrix} \frac{\partial a}{\partial y_1} & \frac{\partial b}{\partial y_1} \\ \cdot & \\ \cdot & \\ \cdot & \\ \frac{\partial a}{\partial y_n} & \frac{\partial b}{\partial y_n} \end{bmatrix} \tag{9}$$

where particular partial derivatives for regression equation coefficients can be calculated from:

$$\frac{\partial a}{\partial y_i}=\frac{x_i-\frac{\Sigma x}{n}}{\Sigma x^2-\frac{(\Sigma x)^2}{n}} \tag{10a}$$

$$\frac{\partial b}{\partial y_i}=\frac{1}{n}-\frac{\Sigma x}{n}\frac{x_i-\frac{\Sigma x}{n}}{\Sigma x^2-\frac{(\Sigma x)^2}{n}} \tag{10b}$$

(sum indexes have been omitted). It leads to the form directly connecting the covariance matrix of regression equation coefficients with measurement results

$$M=\frac{s_0^2}{\Sigma x^2-\frac{\Sigma x}{n}}\begin{bmatrix} 1 & -\frac{\Sigma x}{n} \\ -\frac{\Sigma x}{n} & \frac{\Sigma x^2}{n} \end{bmatrix} \tag{11}$$

Now, it is easy to calculate, necessary for defining ellipse of constant density, the reverse matrix to the covariance matrix M

$$M^{-1}=\frac{1}{s_0^2}\begin{bmatrix} \Sigma x^2 & \Sigma x \\ \Sigma x & n \end{bmatrix} \tag{12}$$

The matrix (12) identifies all confidence ellipses; with the same center point and symmetry axes. The angle φ, between the ellipsis axes and axis of coordinate system can be calculated from the following simple dependence:

$$\mathrm{tg}\varphi=\frac{\Sigma x}{n} \tag{13}$$

The ellipse with constant probability density, determined by value of $C(\alpha)$, has a similar meaning as the confidence interval in the case of one-dimensional random variable. The position of center point defined by the point μ of expected value of two-dimensional distribution is not known a priori.

Repeating many times the estimation of regression equation coefficients, for different sets of experimental results obtained from the few independent experiments, we could expected that evaluated vectors of linear regression line coefficients, will be distributed inside the ellipse, mainly – along to longer axis. It could be presented as the points along the axis, in the form of linear relationship

$$a=\xi b+\delta \tag{14}$$

where ξ and δ are some constants. Values ξ and δ could be calculated from data given by covariance matrix, where constant ξ is equal tgφ.

Interesting things here, are the facts concerning the connection between covariance matrix of equation regression coefficients and the set of experimental results:

- The shape of confidence ellipsis and its position on the plane of regression equation coefficients depends only on adopted experimental plan. It can be determined, with accuracy amounting to constant, before the experiments, with use only data from the plan of experiments.
- One of the essential factors that determine the extension of confidence ellipsis is the assessment of experimental error, expressed as a sum of squares for difference between measuring results for the dependent variable and their assessment expressed in the regression equation. The higher the value s_0^2, the higher evaluation uncertainty and higher the area of two-dimensional confidence interval.
- In numerous cases of experimental measurements performed independently, we should expect a similar shape of confidence ellipse for regression equation coefficients, when the measurements are taken according to similar experimental schedules: e.g. for the same range of temperatures or concentrations.

When determining the vector of regression equation coefficients that define two independently performed experiments, the same question appears as in the case of single-dimensional random variable; whether the observed difference is significant; whether the two results (two points) are significantly different or whether they should be considered to the same population. It signifies the need for assessing the conformability of both results, including the following procedural stages:

- for the vector of difference for both vectors of regression equation coefficients a zero hypothesis is assumed about insignificance of noticed difference,
- a certain significance level defining the rejection of this hypothesis,
- to compare the real assessment result with the expected values
- and to reject this hypothesis when the difference exceeds the acceptable level at the assumed significance level.

Because it if easy to prove that in this case the covariance matrix M_R of sum (or difference) of two vectors $[a_1, b_1]^T$ and $[a_2, b_2]^T$.

$$R = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} = \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} - \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} \tag{15a}$$

is a sum of both covariance matrices

$$M_R = M_1 + M_2 \tag{15b}$$

In terms of calculations it requires:

- determining the covariance matrices for the difference between the two vectors – as a sum of both covariance matrices,
- determining the limit value for the chi-square distribution with two degrees of freedom, for the assumed significance level of, e.g. for α=0.05 limit value χ^2 amounts to 5.99,
- calculating the current value $C(\alpha)$ according to (3) for experimental data, and if the obtained value was lower than the limit one we could adopting the zero hypotheses about insignificance of the observed difference.

Estimation of regression coefficients *vs.* linear compensation law

The kinetics of chemical reactions is usually defined with the product of two functions; function of temperature, and the other – dependent on the quantity of substance x (reacting degree, etc.).

$$\frac{dx}{dt} = f(T)g(x) \tag{16}$$

where the first is commonly defined by the Arrhenius equation

$$f(T) = k_0 \exp\left(-\frac{E}{RT}\right) \tag{17}$$

where T signifies absolute temperature, t – time, E – activation energy, R – gas constant, k_0 – pre-exponential constant.

Basing on experimental values of temperature function the parameters of the Arrhenius equation are usually calculated as the coefficients of linear regression equation

$$\ln[f/T)_i] = \ln k_0 - E\frac{1}{RT_i} \tag{18}$$

where the independent variable is $1/(RT)$. The estimation procedure for constant values consists in determining the values of linear regression, according to the least squares method. The vector $[\ln k_0, E]^T$ confidence ellipse determined on the basis of covariance matrix features main data axis at the angle φ amounting to

$$\mathrm{tg}\varphi = \frac{1}{nR}\sum_{i=1}^{i=n}\frac{1}{T_i} \tag{19}$$

The linear compensation law suggests the existence of a constant linkage between activation energy and the pre-exponential constant for a certain group of substances.

$$\ln k_0 = mE + n \tag{20}$$

The values m and n are constant parameters of the linear compensation equation for the group of experimental data, whose value cannot be theoretically justified. But if this only apparent effect, similar to described by (14), then parameter m should be close to average value of experiment temperature program (19).

Experimental repeatability

Apparently, it is obvious whether the experiment in progress is a repetition of a previous one or it is an entirely independent experiment. If the subject are e.g. the decomposition of some compound that differs only in the type of the cation, then each next experiment describes, according to experimenter's intention, other chemical reality. In particular experiments we obtain different results, which can be interpreted as the impact of the cation type on the behavior of similar compounds.

However, in reality often it does not have to be the case. If the research method is not sensitive enough, the subject of the measurement is an 'average' behavior of compounds (of the group), as the experimental change introduced e.g. the cation has too little impact to make any significant changes in measured values. In some extreme cases, the experimental error might exceed the influence of intentional change made in experimental conditions. The experiments performed in sequence, belonged to the same group, can be considered to be really a repetition of the same experiment. Then the respective estimations of the value pairs: activation energy value and pre-exponential constant would be located inside a specified confidence ellipse inclined at the angle φ, whose value can be assessed according to (19) on the basis of available information on the experiment schedule (preset temperature values). Graphically, it would have the form of points located along some 'average' section. Omitting all more precise static calculations it is possible to say that if the value of determined linear compensation equation m (20) is close to the mean value of measured temperature, as shown in Eq. (19), then we should consider the fact more thoroughly and analyze the impact of accidental and systematic factors on the quality of measurements.

Conclusions

The above presented simple considerations show how the lack of clarity on what calculation procedures used in processing experimental results are, can lead to a logical error. The so-called 'linear compensation law', frequently observed in the case of taking kinetic measurements, is probably only the result of shared action of relatively high measuring error and improper interpretation of character and statistical properties of estimated random values.

The paper includes a simple procedure allowing us to distinguish between a real impact of changed experimental parameters on measurement results and the phenomenon of apparent correlations resulting only from improper usage of standard procedures in processing experimental results. Although, the presented procedure is of some more general importance, it should be applied in all cases when analyzing experiments defined with linear regression equation, as it enables us to create a reliable assessment of measurement accuracy and the impact of experimental variables on experimental results.

Although, the analysis of the incredibility of linear compensation law presented above should be documented with a wider set of experimental results, it is quite obvious that initial assessments [8–11] are enough to show a serious risk concerning the problem. The authors here did not intend to review any experimental papers published by other scientists but they just wanted to share their notes helping to finally solve the problem by any interested experimenter and for his/her own experimental data. The fact that the linear compensation law especially often appears when interpreting the measurement results of chemical process kinetics should make us seriously consider the issue. As the measurements are extremely difficult both in terms of preparing a suitable initial material, clearly described in the way allowing us to repeat any measurement, and in terms of obtaining the results that differentiate those measurements depending on the change in properties of the examined substances, extremely essential here is the relationship between the value of effect resulting from the changes made in respective experiments and the value of measuring error typical for the research method applied.

Repeating the experiment, whether consciously or not, means that the next point determining the values for particular regression equation will be situated in the close vicinity of the previous point, inside the ellipse defining the constant probability density. Because of its elongated shape, in a graphic sense, they generate a line that could be easily called a 'compensation line'. The parameter revealing that the observed connection is physical or chemical in its character is the fact that it is possible to describe the geometrical properties of such a 'compensation' line as early as before starting the measurements, using only the information on expected experimental schedule, but not only after completing the test cycle, basing on the measured numerical values.

The lack of wider set of literature references is intentional. It has been done because the presented issue is a special case being considered almost in any handbook concerning the use of statistical methods for processing experimental results. If there it is prepared in the form of a publication, then it results from the constant return to the issue of compensation law in the papers related to domain of chemical reaction kinetics. The paper includes some significant factors (according to authors), but omitted by experimenters during their statistical analysis of the experimental results.

The above presented considerations show the need for reflection whether the ways of concluding based on the apparently known methods for processing experimental results are justified. Nevertheless, this is not a typical expression in scientific publications though it should be here clearly (but provocatively) stated that in the opinion of authors there are no properly documented cases known for occurring the linear compensation law. All restrictions concerning the above stated conclusion and based on confirmed measurement results will receive a warm welcome.

Acknowledgements

The paper has been prepared within the Polish grant No. 3 T08B 052 28. Authors wish to cordially thank for the possibility of carrying out the paper.

References

1 R. Lopez-Fonseca, I. Landa, M. A. Gutierrez-Ortiz and J. R. Gonzales-Velasco, J. Therm. Anal. Cal., 80 (2005) 65.
2 S. V. Vyazovkin and A. I. Lesnikovich, Thermochim. Acta, 128 (1988) 69.
3 B. Jankovic, B. Adnadevic and J. Jovanovic, J. Therm. Anal. Cal., 82 (2005) 7.
4 P. D. Garn, J. Thermal Anal., 7 (1975) 475.
5 J. Zsakó and M. Lungu, J. Thermal Anal., 5 (1973) 77.
6 D. Klinar, J. Golob and M. Krajnc, Chem. Biochem. Eng. Q., 18 (2004) 65.
7 X. G. Li and M. R. Huang, Polym. Degrad. Stab., 64 (1999) 81.
8 Z. Kolenda, J. Norwisz and N. Hajduk, Proceeding of the VI International Conference on Thermal Analysis, (1980), Vol. 1, p. 4145.
9 Z. Smieszek. Z. Kolenda, J. Norwisz and N. Hajduk, J. Thermal Anal., 25 (1982) 377.
10 J. Norwisz, Z. Smieszek and Z. Kolenda, Thermochim. Acta, 156 (1989) 313.
11 J. Norwisz, Z. Smieszek and Z. Kolenda, Thermochim. Acta, 156 (1989) 321.
12 J. M. Thomas and W. J. Thomas, Introduction to the Principles of Heterogenous Catalysis, Acad. Press, New York, London 1967, p. 263.
13 P. D. Garn, J. Thermal Anal., 7 (1975) 475.
14 S. V. Vyazovkin and W. Linert, J. Solid State Chem., 114 (1995) 392.
15 D. F. Morrison, Multivariate Statistical Methods, McGraw-Hill Book Company, 1976.
16 A. D. Aczel, Complete Business Statistics, Irwin, Burr Ridge, Illinois, Boston Massachusetts, Sydney 1993.

DOI: 10.1007/s10973-006-8139-4