

STATISTICAL MODEL OF SYSTEMATIC ERRORS: LINEAR ERROR MODEL

E.B. Rudnyi

Department of Chemistry
Moscow State University
119899 Moscow, Russia
e-mail rudnyi@comp.chem.msu.su
home pages <http://www.chem.msu.su/~rudnyi/welcome.html>

ABSTRACT

A statistical model of systematic errors for processing results of a few experiments is presented. The model includes the shift and tilt systematic errors (the linear error model). To obtain estimates of unknown parameters and unknown variance components, the maximum likelihood method is suggested to apply. The algorithm to maximize a likelihood function in the case of a non-linear physico-chemical model and the linear error model is described. The method has been applied to both simulated and real data (the vaporization of KCl). The results obtained are compared with those of the least squares method.

LIST OF SYMBOLS

| | |
|----------------------|--|
| y_{ij} | experimental observation |
| Θ | vector of unknown parameters |
| $i = 1, \dots, M$ | index enumerating series |
| M | number of all the series |
| $j = 1, \dots, N_i$ | index enumerating points in the i -th series |
| N_i | number of points in the i -th series |
| SS | general sum of squared deviations |
| SS_i | sum of squared deviations of the i -th experiment |
| L | function, the maximum of which coincides with the maximum of the likelihood function |
| ε_{ij} | total error in y_{ij} |
| ε | vector of all the deviates ε_{ij} , the number of elements is $\sum_i N_i$ |
| ε_i | vector of the deviates in the i -th series, the number of elements is N_i |
| $\varepsilon_{r,ij}$ | reproducibility error in the y_{ij} |
| $\varepsilon_{a,i}$ | shift systematic error in the i -th series |
| $\varepsilon_{b,i}$ | tilt systematic error in the i -th series |

| | |
|--|--|
| $\sigma_{r,i}^2$ | variance of reproducibility errors in the i -th series |
| $\sigma_{a,i}^2$ | variance of shift systematic error in the i -th series |
| $\sigma_{b,i}^2$ | variance of tilt systematic error in the i -th series |
| $\gamma_{a,i} = \sigma_{a,i}^2 / \sigma_{r,i}^2$ | |
| $\gamma_{b,i} = \sigma_{b,i}^2 / \sigma_{r,i}^2$ | |
| $\delta_{a,i} = \{1 - (1 + N_i \gamma_{a,i})^{-1/2}\} / N_i$ | |
| $\delta_{b,i} = \{1 - (1 + P_i \gamma_{b,i})^{-1/2}\} / P_i$ | |
| D | dispersion matrix |
| I_i | identity matrix, dimension is N_i |
| V_i | dispersion matrix of the i -th series |
| 1_i | vector composed from ones, the number of elements is N_i |
| x_i | vector composed from $(x_{ij} - x_i)$ |
| $x_i = (\sum_j x_{ij}) / N_i$ | mean of x_{ij} in the i -th series |
| $P_i = \sum_j (x_{ij} - x_i)^2$ | |

CONTENTS

1. Introduction
2. Notation and conventions
 - 2.1. Physico-chemical model
 - 2.2. Statistical model of reproducibility and systematic errors
 - 2.3. Estimation of variance components
3. Simulated data - the same variances of errors
 - 3.1. One-way classification
 - 3.2. Linear regression
4. Real case (vaporization of KCl) - different variances of errors
 - 4.1. Vaporization thermodynamics of KCl
 - 4.2. Analysis of deviates
 - 4.3. Expert analysis
5. Comparison with the least squares method
6. Algorithm
 - 6.1. Some expressions for the dispersion matrix in the closed form
 - 6.2. Finding the maximum of the likelihood function
7. Acknowledgment
8. References

1. INTRODUCTION

The difference among values obtained in several experiments (between-errors) are usually greater than reproducibility scatter within an experiment (within-errors). In this case, it is said that there are experimental systematic errors, and the results measured in one series disperse over the line somewhat shifted from the "true" equation.

To obtain a first impression of the task in question, have a look over Fig. 1 and 2. Values presented simulate results of eight experiments in Fig. 1 and six experiments in Fig. 2 when systematic errors increase from case *a* to *d*. In case *a* there are no systematic errors, and case *d* displays rather extreme case when the between-errors are much greater than the within-errors. The figures will be explained and discussed in section 3.

To take into account systematic errors, the present work employs the special formalism of mathematical statistics called estimation of variance components [1-5]. It is assumed that there are results of a few experiments (6 and up) available. The more number of experiments, the more reliable results are obtained.

2. NOTATION AND CONVENTIONS

2.1. Physico-chemical model

The results of an experiment will be called a *series*. A series is a set of experimental points $\{y_{ij}, x_{ij}\}$ where y_{ij} is usually what is measured and x_{ij} is what is set. Subscript i stands for the ordinal number of the series ($i = 1, \dots, M$); M is a total number of the series. Subscript j enumerates the experimental observations in the i -th series ($j = 1, \dots, N_i$); N_i is a number of the points in the i -th series.

Some *a priori* known relationship is assumed to be between y and x that will be called as a physico-chemical model (this term has been preferred instead of mathematical model because it seems to be little less abstract)

$$y = f(x; \Theta)$$

and that contains several unknown parameters, vector Θ . In the general case, the dependence of y from parameters in vector Θ is non-linear. Note, that different equations can be used for different series provided the equations contain the same unknown parameters.

The goal is to infer estimates and their variances of the unknown parameters in the physico-chemical model. To this end, there are $\sum_i N_i$ equations but because of the measurement errors they are not compatible and have no solution. It means that while real experimental values, y_{ij} and x_{ij} are substituted in the physico-chemical model it should be written as

$$y_{ij} = f_i(x_{ij}; \Theta) + \varepsilon_{ij} \quad (1)$$

Notation f_i underlines that physico-chemical models may be different for different series.

The system of equations (1) is already feasible but has many solutions because the total number of unknowns including measurement errors is greater than the number of equations. Thus, the "best" solution is to be chosen.

If mere estimates of parameters are necessary, any criterion according one's aesthetic preferences (the least sum of squared deviations, the least sum of modules, the criterion of minimax, etc.) can be employed. The meaningful problem appears if reliability of parameters inferred and/or reliability of the physico-chemical model is needed. To this end, there is no other means but mathematical statistics. The first necessary step on this way is to choose an appropriate error model.

2.2. Statistical model of reproducibility and systematic errors

Mathematical statistics suggests us to consider measurement errors as the realizations of a random quantity. Metaphorically speaking, the random quantity is a lottery box on a front panel of which some values appear in turn - they will be errors. Nature takes the role of a contest master. It picks an exact value of a physico-chemical quantity (don't ask me where it resides), reads a current error on the lottery box and adds the latter to the former. We become players in the contest when only final results are available (through our apparatus) and we should make guesses on true values of the physico-chemical quantity.

If nothing can be said about the lottery box - nothing can be said at all. Let us make general conjectures counting on that even if Nature plays dice it does not cheat and that the form of the physico-chemical model has been specified correctly by a scientist.

First, an expected value of errors (a mean of all the elements in the lottery box) is zero

$$M(\varepsilon_{ij}) = 0$$

Second, the error distribution is close to normal (Gauss distribution) or, at least, the variance of the distribution is finite.

Third, the lottery box is the same for all the experimental points in one series although Nature may use different lottery boxes for different experiments. It means that the variances in a series are homogeneous

$$D(\varepsilon_{ij}) = \sigma_i^2$$

To simulate systematic errors, let us start with the calibration errors (for example, see Fig. 1). One can say that a total error ε_{ij} is the sum of two errors - a reproducibility error $\varepsilon_{r,ij}$ and a shift systematic error $\varepsilon_{a,i}$

$$\varepsilon_{ij} = \varepsilon_{r,ij} + \varepsilon_{a,i}$$

The metaphor is extended. Here Nature already uses two lottery boxes for a series. Before measurements in the i -th series Nature takes a value from a lottery box of shift systematic errors and employs it as the base for all the reproducibility errors (to be read from the reproducibility lottery box) in this series. Such a model can explain a shift of all the values in the i -th series from true ones due to the calibration error.

In this paper, the distribution of systematic errors is assumed to be close to the Gauss distribution. An expected value of systematic errors (a mean of all the elements in the lottery box of systematic errors) is zero, and the variance is finite. Thus, an expected value of a total error is still zero

$$M(\varepsilon_{ij}) = M(\varepsilon_{r,ij}) + M(\varepsilon_{a,i}) = 0$$

The main difference after introducing shift systematic errors is that the total errors in a series will be correlated with each other (provided that reproducibility and systematic errors are non-correlated)

$$\text{cov}(\varepsilon_{ij}, \varepsilon_{kl}) = \sigma_{a,i}^2$$

where $\sigma_{a,i}^2$ is a variance of the shift systematic error. As a result, a dispersion matrix of errors changes its form from diagonal to block-diagonal (see below).

The systematic shift error only is not enough for the adequate description of real data. Let us continue and say that the systematic error is a function of x . The simplest relationship is the linear one

$$\varepsilon_{ij} = \varepsilon_{r,ij} + \varepsilon_{a,i} + \varepsilon_{b,i}(x_{ij} - x_i) \quad (2)$$

where

$$x_i = (\sum_j x_{ij})/N_i$$

Eq. (2) will be referenced as the linear error model (note that this is a special case of the general error model described in Refs [1 - 5]). It simulates both shift and tilt systematic errors. The latter describes the case when apparatus gives a somewhat wrong functional dependence. In the framework of metaphor, Nature should use three lottery boxes (one for reproducibility and two for systematic errors) for each series.

In many cases, Eq. (2) is enough to represent behavior of systematic errors in both linear and non-linear physico-chemical models. In the latter case, Eq. (2) can be seen as the expansion in series of the systematic error function. The Eq. (2) accounts for the main effects of systematic errors, and introducing new terms not only complicates the computing procedure but also gives badly identified variance components.

Once more it should be mentioned that Eq. (2) does not mean the linearization of the physico-chemical model (Eq. 1 has not been touched) but the linearization of the systematic error function.

It is handy to discuss the systematic errors introduced above by employing two quantities

$$\gamma_{a,i} = \sigma_{a,i}^2 / \sigma_{r,i}^2, \quad \gamma_{b,i} = \sigma_{b,i}^2 / \sigma_{r,i}^2,$$

where $\sigma_{r,i}^2$, $\sigma_{a,i}^2$ and $\sigma_{b,i}^2$ are the variances of the reproducibility error and the shift and tilt systematic errors accordingly.

2.3. Estimation of variance components

There is a set of experimental points $\{y_{ij}, x_{ij}\}$ obtained in M experiments. The correct physico-chemical model (Eq. 1) is chosen and the linear error model (2) is accepted. It is necessary to infer estimates of the unknown parameters (vector Θ) and the variance components (variances of reproducibility errors $\sigma_{r,i}^2$, shift $\sigma_{a,i}^2$ and tilt $\sigma_{b,i}^2$ systematic errors).

To solve the task one must start with some additional hypotheses about the relation between variances in different experiments - were or were not the lottery boxes the same? This question will be discussed later in section 3 and 4.

The error model (2) brings forth the block-diagonal dispersion matrix of the total errors

$$\mathbf{D}(\boldsymbol{\varepsilon}) = \text{diag}\{\mathbf{V}_i\} \quad (3)$$

where $\boldsymbol{\varepsilon}$ is a vector of the total errors ε_{ij} (the number of the elements is $\sum_i N_i$), \mathbf{V}_i is the dispersion matrix of the i -th series

$$\mathbf{V}_i = \sigma_{r,i}^2 \mathbf{I}_i + \sigma_{a,i}^2 \mathbf{1}_i \mathbf{1}_i' + \sigma_{b,i}^2 \mathbf{x}_i \mathbf{x}_i' \quad (4)$$

where, in turn, \mathbf{I}_i and \mathbf{x}_i are vectors with the number of elements equal to N_i , the former comprises units, the latter does quantities ($x_{ij} - x_i$). While Eq. (4) is obtained the reproducibility and systematic errors are assumed to be non-correlated.

Much attention has been paid to the task described above recently [1 - 5]. From a number of methods suggested the maximum likelihood method was chosen. The main reason for that choice was that its simple and effective numerical implementation for the case under study happened to have been found. The maximum likelihood method gives some downward bias in estimated variance components [1, 2]. A value of the bias is about amount of p/n where p is the number of unknown parameters and n is the number of the experimental points. In cases, when $p \gg n$ the bias can be neglected. The restricted maximum likelihood method [1, 2] is free from that drawback but its numerical implementation is more complicated.

Maximizing the likelihood function gives estimates of the parameters as well as the variance components. Provided all the errors are described by multinormal distribution the maximum of the likelihood function coincides with the maximum of

$$L = -\ln \{\det[\mathbf{D}(\boldsymbol{\varepsilon})]\} - \boldsymbol{\varepsilon}' \mathbf{D}(\boldsymbol{\varepsilon})^{-1} \boldsymbol{\varepsilon} \quad (5)$$

If the error distribution is unknown, finding the maximum of function (5) may be viewed as a heuristic procedure that gives not the worst estimates of the physico-chemical parameters and the variance components.

Note that the least squares method is a special case of maximizing Eq. (5) when the variance components are known *a priori*. More exactly, the variance matrix should be supplied up to an unknown constant that can be written in terms of the weighted matrix

$$\mathbf{W} = k \mathbf{D}(\boldsymbol{\varepsilon})^{-1} \quad (6)$$

In this case, the maximum of (5) occurs at the same time as the minimum of the generalized sum of squared deviations

$$SS = \boldsymbol{\varepsilon}' \mathbf{W} \boldsymbol{\varepsilon} \quad (7)$$

Actually, the algorithm for finding the maximum of function (5) in the case of a non-linear physico-chemical model with the linear error model (see section 6) can be roughly seen as an "iterative" least squares method. This matters will be discussed in Section 5.

The variances of the parameters in vector $\boldsymbol{\Theta}$ can be estimated as usual

$$\mathbf{D}(\boldsymbol{\Theta}) = \{\mathbf{A}' \mathbf{D}(\boldsymbol{\varepsilon})^{-1} \mathbf{A}\}^{-1} \quad (8)$$

where the variance component estimates found after minimizing of function (5) are substituted in the dispersion matrix $\mathbf{D}(\varepsilon)$, and the matrix $\mathbf{A} = \{y_{ij}/\theta_j\}$ is the Jacobean estimated at final values of parameters.

3. SIMULATED DATA - THE SAME VARIANCES OF ERRORS

What may the following mean - systematic errors are the same? If it means that the values of the systematic errors are the same in all the experiments ($\varepsilon_{a,i} = \text{const}$ and $\varepsilon_{b,i} = \text{const}$), we would have the case quite similar with the one when there are no systematic errors at all.

The clause "systematic errors are the same" will get better sense if it means the same distribution of the systematic errors in all the series ($\sigma_{a,i}^2 = \text{const} = \sigma_a^2$ and $\sigma_{b,i}^2 = \text{const} = \sigma_b^2$). Here the instances of the systematic errors will be different and one can obtain estimates of variances of both reproducibility and systematic errors.

If the assumption $\sigma_{r,i}^2 = \text{const} = \sigma_r^2$ is added the simplest case is obtained. There are only three unknown variance components in all the experiments in addition to the unknown physico-chemical parameters. Under the metaphor of section 2.2 it means that Nature has employed three lottery boxes for all the series. The current section is devoted to that particular case.

The method can be checked if it applies to the task where the answer is known. In mathematical statistics, this supposes using of a pseudo-random generator for preparing the simulated results. On this way, two examples based on the normal distribution have been

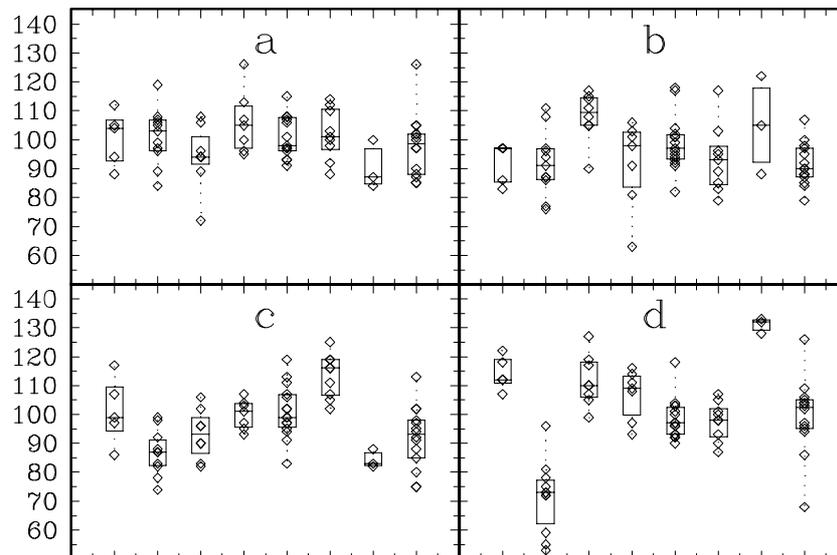


Fig. 1. Simulated values of eight series for one-way classification (Eq. 9). The variance of shift systematic errors increases from case *a* to *d* (see Table 1). The box contains half of all the experimental values in a series, and the middle of the box shows up the median.

made.

To facilitate calculations, two MS-DOS utilities have been written [6] (they are available over Internet). The GENERATE generates pseudo-experimental values with the normal distribution, and the LINEAR maximizes the function (5) for the models considered in this section.

3.1. One-way classification

The simplest physico-chemical model is connected with measuring the same quantity in a few experiments

$$y_{ij} = \mathbf{a} + \varepsilon_{r,ij} + \varepsilon_{a,i} \quad (9)$$

Traditionally, it is called by one-way classification. In physical chemistry, thermochemistry can be taken as an example: "The enthalpy of formation of A has been measured in several laboratories. Estimate the enthalpy of formation."

Fig. 1 displays the pseudo-experimental values of some quantity measured in eight experiments with the expected value of $\mathbf{a} = 100$ and the standard deviation of reproducibility of $\sigma_r = 10$. Two distinct normal distributions were used for all the eight series: $\sigma_{r,i}^2 = \sigma_r^2$ and $\sigma_{a,i}^2 = \sigma_a^2$ (tilt systematic errors are not applicable here). Four cases from a to d differ by the value of $\gamma_a = \sigma_a^2/\sigma_r^2$ (see Table 1). Case a simulates the results of eight experiments without systematic errors, and the cases from b to d correspond to shift systematic errors being on the

Table 1. The initial parameter and variance values for the one-way classification simulation and the results obtained by processing the pseudo-experimental values.

| | \mathbf{a} | $\sigma(\mathbf{a})$ | σ_r | $\sqrt{\gamma_a}$ | \mathbf{a} | $\sigma(\mathbf{a})$ | σ_r | $\sqrt{\gamma_a}$ |
|------|--------------|----------------------|------------|-------------------|--------------|----------------------|------------|-------------------|
| | a | | | | b | | | |
| true | 100.0 | 1.18 | 10.0 | 0.00 | 100.0 | 2.18 | 10.0 | 0.50 |
| ML | 99.8 | 1.15 | 9.7 | 0.00 | 96.1 | 2.09 | 10.4 | 0.44 |
| OLS | 99.8 | 1.15 | 9.7 | 0.00 | 95.8 | 1.33 | 11.3 | 0.00 |
| | c | | | | d | | | |
| true | 100.0 | 3.77 | 10.0 | 1.00 | 100.0 | 7.20 | 10.0 | 2.00 |
| ML | 96.7 | 3.05 | 8.8 | 0.91 | 103.5 | 5.57 | 9.6 | 1.60 |
| OLS | 97.0 | 1.38 | 11.7 | 0.00 | 99.2 | 1.97 | 16.7 | 0.00 |

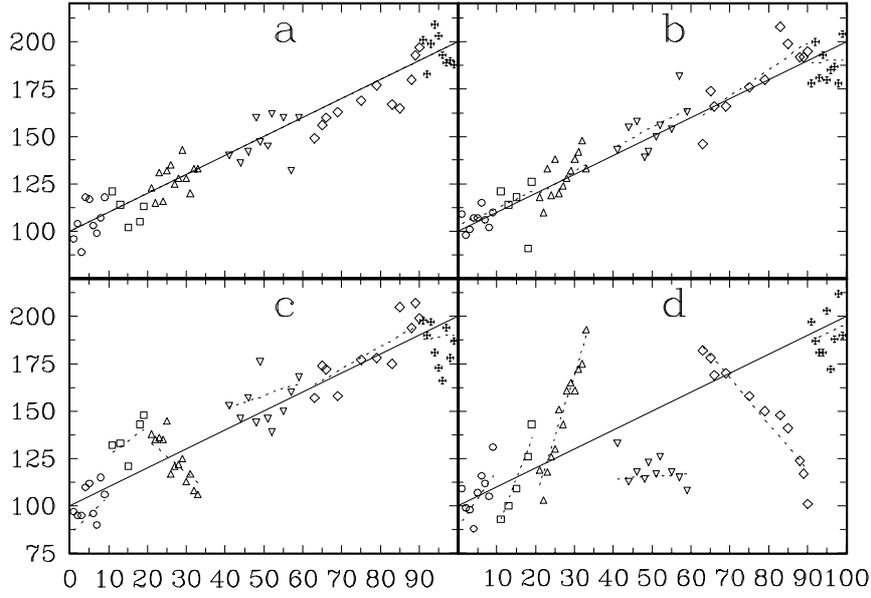


Fig. 2. Simulated values of six series for the simple linear model (10). The variance of shift and tilt systematic errors increases from case *a* to *d* (see Table 2).

level of 50%, 100%, and 200% from reproducibility errors accordingly.

The initial values employed for the simulation, and the results of processing the pseudo-experimental values by means of maximizing the likelihood function (ML) and also by the ordinary least squares (OLS) method are given in Table 1. It can be seen, that the ML results, the parameter and variance estimates, are quite close to the true values. The OLS results will be discussed in Section 5.

3.2. Linear regression

Quite often, it is necessary to estimate parameters of a straight line from results of several experiments. Altogether with the linear error model, the problem can be written as follows

$$y_{ij} = a + b x_{ij} + \varepsilon_{r,ij} + \varepsilon_{a,i} + \varepsilon_{b,i}(x_{ij} - x_i) \quad (10)$$

Fig. 2 shows pseudo-experimental values of six experiments (equation is $y = 100 + x$, the standard deviation of reproducibility $\sigma_r = 10$) when it was set that $\sigma_{r,i}^2 = \text{const} = \sigma_r^2$, $\sigma_{a,i}^2 = \text{const} = \sigma_a^2$ and $\sigma_{b,i}^2 = \text{const} = \sigma_b^2$ (the same three lottery boxes for all the experiments). Like in the one-way simulation, the level of the systematic errors (values of $\gamma_a = \sigma_a^2/\sigma_r^2$ and $\gamma_b = \sigma_b^2/\sigma_r^2$) increases from case *a* to *d* (see Table 2).

Table 2 lists the initial values and the results of processing pseudo-experimental values by means of maximizing the likelihood function (ML) and by the ordinary least squares (OLS) method. Again, the ML results, the parameter and variance estimates, are quite close to the true values, and the OLS results will be discussed in Section 5.

Table 2. The initial parameter and variance values for the linear regression simulation and the results obtained by processing the pseudo-experimental values.

| | a | $\sigma(a)$ | b | $\sigma(b)$ | σ_r | $\sqrt{\gamma_a}$ | $\sqrt{\gamma_b}$ |
|----------|-------|-------------|-------|-------------|------------|-------------------|-------------------|
| <i>a</i> | | | | | | | |
| true | 100.0 | 2.35 | 1.000 | 0.041 | 10.0 | 0.00 | 0.00 |
| ML | 100.0 | 2.09 | 0.963 | 0.037 | 8.9 | 0.00 | 0.00 |
| OLS | 100.0 | 2.09 | 0.963 | 0.037 | 8.9 | 0.00 | 0.00 |
| <i>b</i> | | | | | | | |
| true | 100.0 | 4.20 | 1.000 | 0.074 | 10.0 | 0.50 | 0.05 |
| ML | 102.6 | 2.67 | 0.959 | 0.047 | 9.3 | 0.23 | 0.04 |
| OLS | 102.5 | 2.30 | 0.965 | 0.040 | 9.8 | 0.00 | 0.00 |
| <i>c</i> | | | | | | | |
| true | 100.0 | 7.25 | 1.000 | 0.129 | 10.0 | 1.00 | 0.10 |
| ML | 107.4 | 5.73 | 0.873 | 0.102 | 9.4 | 0.80 | 0.18 |
| OLS | 104.3 | 3.172 | 0.920 | 0.056 | 13.5 | 0.00 | 0.00 |
| <i>d</i> | | | | | | | |
| true | 100.0 | 13.79 | 1.000 | 0.245 | 10.0 | 2.00 | 0.20 |
| ML | 104.2 | 10.92 | 0.747 | 0.196 | 8.9 | 1.73 | 0.39 |
| OLS | 110.2 | 6.06 | 0.635 | 0.106 | 25.8 | 0.00 | 0.00 |

Once more, you can generate your own sets of pseudo-experimental values for models (9) and (10) by means of the utility GENERATE and to process them with the utility LINEAR [6].

4. REAL CASE (VAPORIZATION OF KCL) - DIFFERENT VARIANCES OF ERRORS

The hypothesis that the corresponding variance components are the same in all the experiments (see section 3) does not seem to be adequate to describe the real world. More sensible if one will expect that, as it was stated in section 1, Nature utilized different lottery boxes for different series. Such a hypothesis that all variance components, $\sigma_{r,i}^2$, $\sigma_{a,i}^2$ and $\sigma_{b,i}^2$ are all different in distinct series is diametrically opposite to the assumption of section 3 and means that there are $3*M$ (M is a number of the series) unknown variance components besides

the unknown parameters in the physico-chemical model. Unfortunately, this postulate is too broad to be applicable in the practical work.

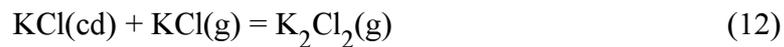
The problem is not in the number of the unknown variance components by itself - all of them, in principle, are identifiable. The problem is that there are many distinct solutions that can not be discriminated without additional *a priori* information. In the maximum likelihood method, it means that function (5) possesses many local maximums which differ insignificantly by value of L. My experience shows that the situation is as follows. Each maximum corresponds to the case when zero estimates of systematic errors are ascribed to a particular series and the estimates of others systematic errors are non-zero - the fitting curve in this case goes through that particular series and all the blame is on other experiments. Thus under the hypothesis stated above, we would have a number of solutions each of which corresponded to *a priori* information that some experiment is right and all others are wrong.

To cope with this situation, it is necessary to find a middle ground between two extreme cases. One should separate the series into some groups and assume that the matching variance components are the same inside a group. Pragmatically speaking, this should leave only one global maximum at the likelihood function (5) and hence one solution only.

A stratagem as follows can be suggested. Start with the hypothesis that the analogous variances are same, $\sigma_{r,i}^2 = \sigma_r^2$, $\sigma_{a,i}^2 = \sigma_a^2$ and $\sigma_{b,i}^2 = \sigma_b^2$. Then after both the analysis of deviates and *a priori* information based upon an expert examination, choose the most appropriate hypothesis. Such an approach is described in this section with example of the real data pertained to the vaporization thermodynamics of potassium chloride [7].

4.1. Vaporization thermodynamics of KCl

In the vapors of potassium chloride, there are monomer KCl and dimer K_2Cl_2 molecules. The equilibria in the saturated vapors can be described by the two reactions



equilibrium constants of which depend on unknown enthalpies and entropies

$$\ln K^0(11) = \ln p_m = - \Delta_r H_m^0 / RT + \Delta_r S_m^0 / R \quad (13)$$

$$\ln K^0(12) = \ln (p_d / p_m) = - \Delta_r H_{dm}^0 / RT + \Delta_r S_{dm}^0 / R \quad (14)$$

where p_m is a partial pressure of monomers, and p_d is that of dimers. There are the experimental values for the total pressure $p_{tot,ij}$, the effective pressure of the Knudsen effusion method $p_{KE,ij}$ and the effective pressure of the transpiration method $p_{TR,ij}$

$$\ln p_{tot,ij} = \ln p_m + \ln(1 + p_d / p_m) + \varepsilon_{ij} \quad (15)$$

$$\ln p_{KE,ij} = \ln p_m + \ln(1 + \sqrt{2} p_d / p_m) + \varepsilon_{ij} \quad (16)$$

$$\ln p_{TR,ij} = \ln p_m + \ln(1 + 2 p_d/p_m) - \ln\{1 + p_m (p_d/p_m)/p_{sys}\} + \varepsilon_{ij} \quad (17)$$

in the temperature interval from 624 to 1691 K. Overall, there are 460 experimental point measured in 29 experiments by different authors. p_{sys} in Eq. (17) stands for the total system pressure in the transpiration apparatus.

The enthalpies and entropies of reactions (11) and (12) depend on temperature by themselves, but their temperature function is known - there are rather reliable values of the heat capacity for condensed potassium chloride, monomer and dimer molecules. As a result, the values of $\Delta_r H_m^\circ$, $\Delta_r S_m^\circ$, $\Delta_r H_{dm}^\circ$, and $\Delta_r S_{dm}^\circ$ at 1044 K (melting temperature) for the sublimation process have been chosen as the unknowns, and necessary corrections were made at other temperatures.

Equations from (13) to (17) make up the physico-chemical model (compare with Eq. 1) where $\ln p_{ij}$ stands for y_{ij} , $1/T_{ij}$ for x_{ij} , and the enthalpies and entropies, $\Delta_r H_m^\circ$, $\Delta_r S_m^\circ$, $\Delta_r H_{dm}^\circ$, $\Delta_r S_{dm}^\circ$ compose the vector Θ . More details on thermodynamics background, the references to the original experimental papers and the discussion of the results obtained from physical chemistry point of view are elsewhere [7].

Figures from 3 to 5 show the experimental points for the total pressure, Knudsen effusion and transpiration methods. To enhance the scale, the total deviates

$$\varepsilon_{ij} = \ln p_{ij} - \ln p_{ij}^{calc}$$

are given where p_{ij} stands for any of $p_{tot,ij}$, $p_{KE,ij}$, $p_{TR,ij}$ and p_{ij}^{calc} is calculated for the recommended set of enthalpies and entropies (solution ML, see below).

First, it can be said that the difference among distinct work is not that big, 10-15% of the absolute pressure value. The latter changes by itself by ten orders of magnitude from 600 to 1700 K. Second, we clearly have a situation when between-errors (10-15%) are bigger than within-errors (3-5%). Third, the linear error model (2) seems to explain the deviates behavior rather well.

To proceed with the treatment, it is necessary to formulate a statement whether variance components $\sigma_{r,i}^2$, $\sigma_{a,i}^2$ and $\sigma_{b,i}^2$ are the same or not in different experiments. As mentioned above, it can be done by a combination of the analysis of deviates and the expert examination.

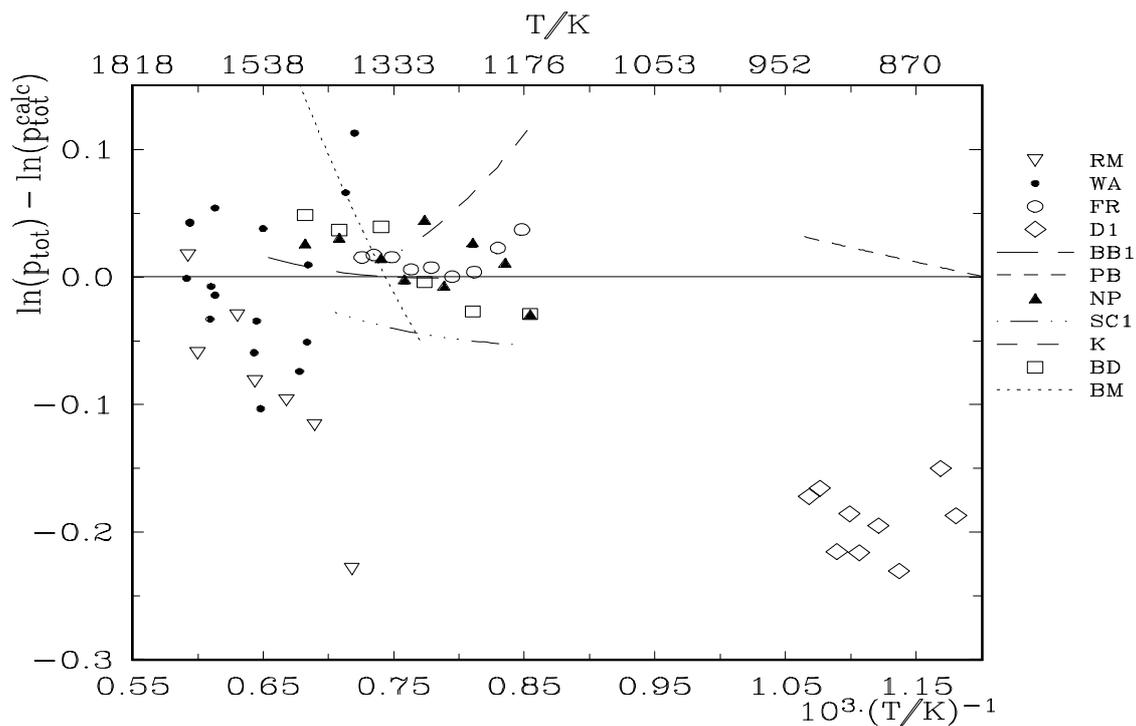


Fig. 3. The deviates of the experimental total pressure values from the recommended solution (ML).

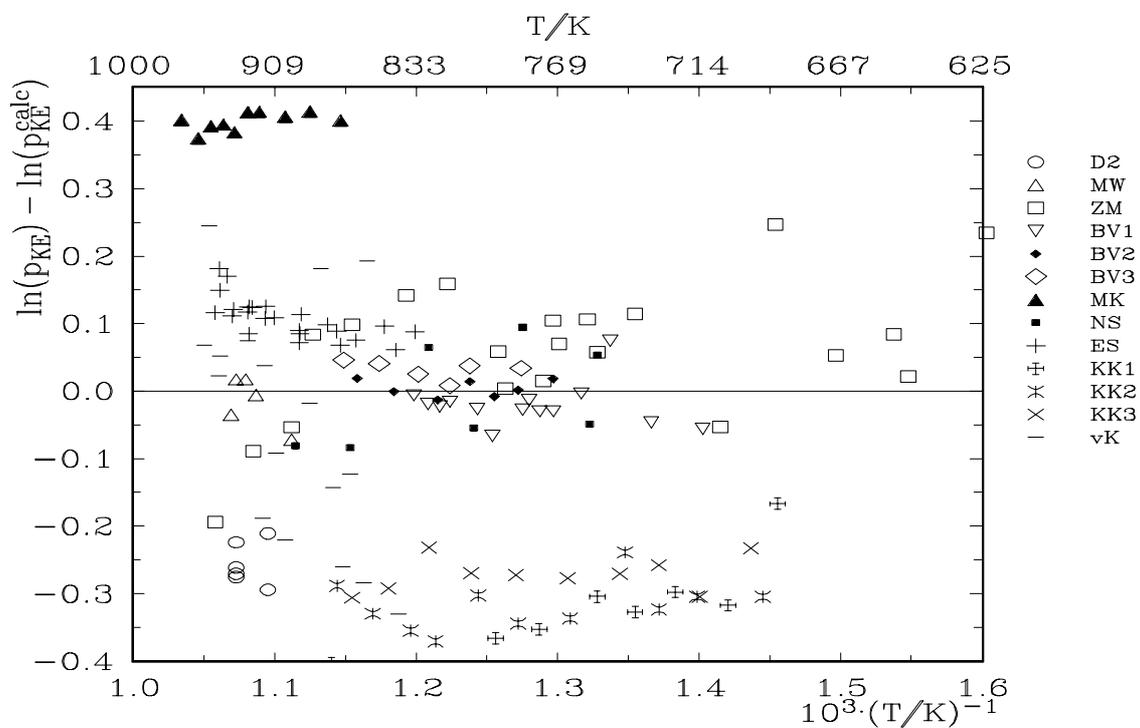


Fig. 4. The deviates of the experimental Knudsen effusion effective pressure values from the recommended solution (ML).

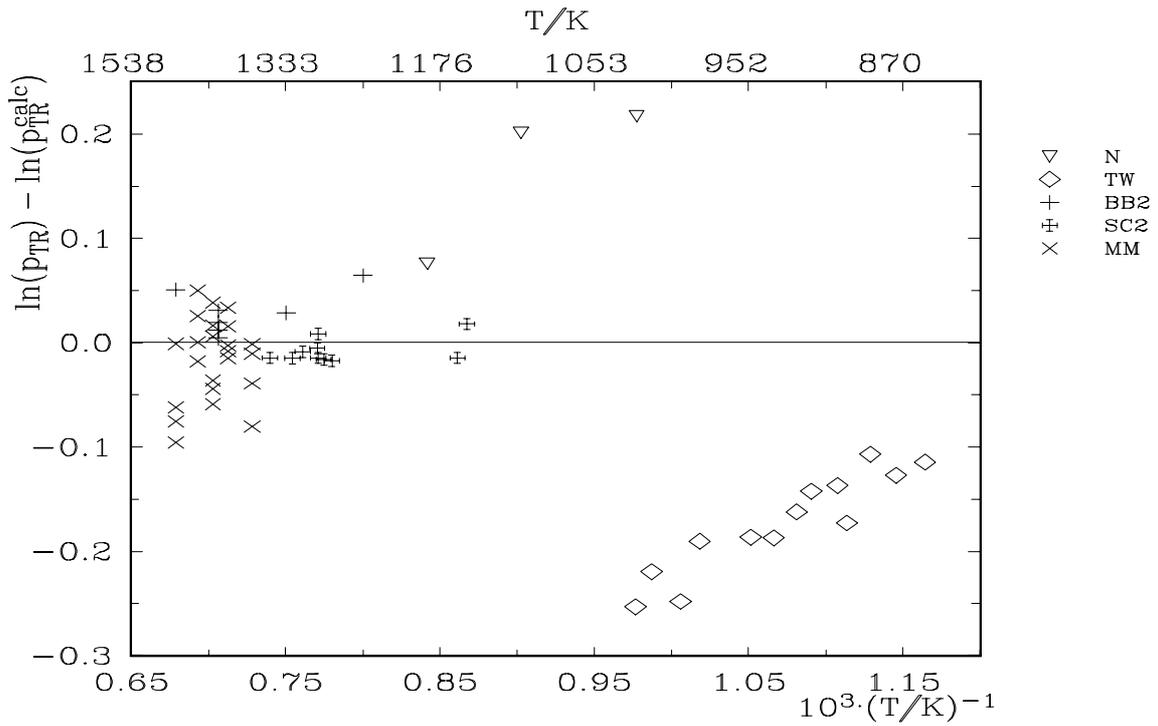


Fig. 5. The deviates of the experimental transpiration effective pressure values from the recommended solution (ML).

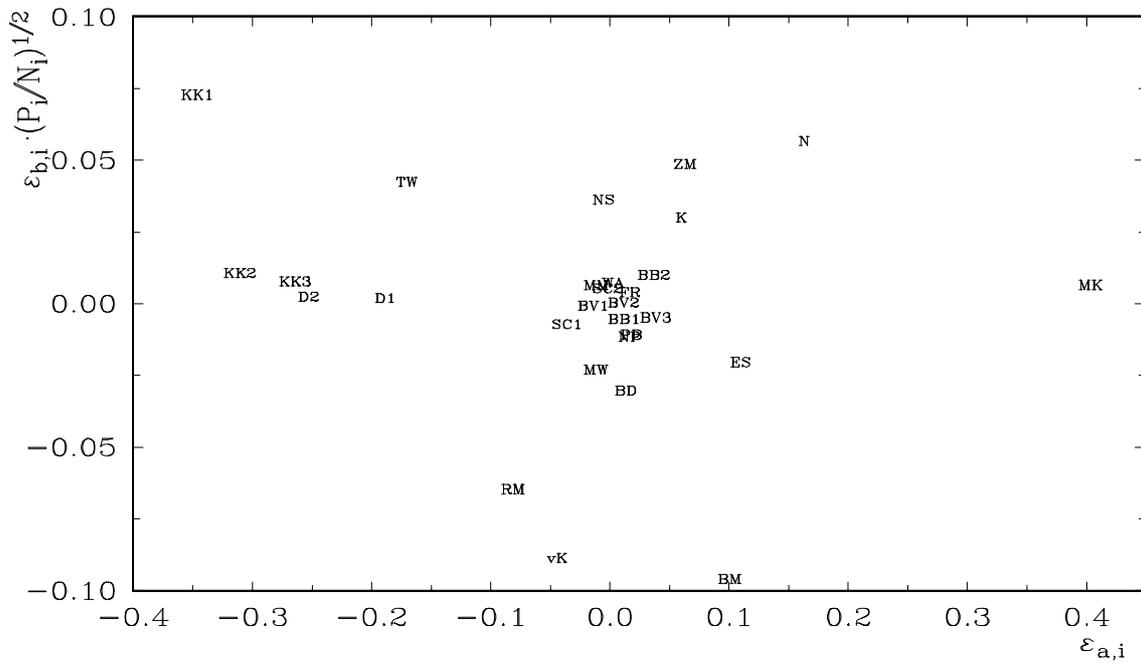


Fig. 6. Tilt systematic errors vs. shift ones. The code of an experiment is used as the mark

4.2. Analysis of deviates

For any set of the parameters of the physico-chemical model it is possible to compute the total deviates

$$\varepsilon_{ij} = y_{ij} - f_i(x_{ij}; \Theta)$$

Under the linear error model (2), the instances of systematic errors $\varepsilon_{a,i}$ and $\varepsilon_{b,i}$ can be estimated by assuming that the sum of reproducibility errors within a series

$$\sum_j \varepsilon_{r,ij}^2 = \sum_j \{\varepsilon_{ij} - \varepsilon_{a,i} - \varepsilon_{b,i}(x_{ij} - x_i)\}^2$$

should be minimal. The minimization gives

$$\varepsilon_{a,i} = (\sum_j \varepsilon_{ij})/N_i$$

$$\varepsilon_{b,i} = \{\sum_j \varepsilon_{ij}(x_{ij} - x_i)\}/P_i$$

where

$$P_i = \sum_j (x_{ij} - x_i)^2$$

Useful is the following relationship

$$(\sum_j \varepsilon_{r,ij}^2)_{\min} = \sum_j \varepsilon_{ij}^2 - N_i \varepsilon_{a,i}^2 - P_i \varepsilon_{b,i}^2$$

which allows to compute the minimal sum of reproducibility errors by one computation cycle. In the case of potassium chloride, the values of $1/T_{ij}$ should be substituted instead of x_{ij} .

Then, a graph when each series presented as a point in axes $\varepsilon_{b,i} - \varepsilon_{a,i}$ is helpful. On this way, fig. 6 is obtained. It displays estimates of the tilt systematic errors vs. the shift ones for 29 experiments on vaporization of KCl [7]. Worthy of noting that no statistical hypotheses were employed, the suggested graph is just another convenient way to show off what agreement among different experiments looks like in addition to fig. 3-5.

The quantities $\varepsilon_{a,i}$ and $\varepsilon_{b,i}$ have different dimensions and this does not allow us to compare their numerical values directly. From this point of view, it is better to employ as the ordinate not $\varepsilon_{b,i}$ but the product $\varepsilon_{b,i}(P_i/N_i)^{1/2}$ (so does fig. 6). The product has the same dimension and sense as the quantity $\varepsilon_{a,i}$. The quantities $\varepsilon_{a,i}$ and $\varepsilon_{b,i}(P_i/N_i)^{1/2}$ display the mean difference between total ε_{ij} and reproducibility $\varepsilon_{r,ij}$ errors due to the shift and the tilt accordingly in a given experiment.

4.3. Expert analysis

Figure 6 indicates that the hypothesis about the alike variance components being the same in all the experiments, $\sigma_{a,i}^2 = const = \sigma_a^2$ and $\sigma_{b,i}^2 = const = \sigma_b^2$ is not appropriate. This is in accordance with the expert expectation, because the 29 works on the vaporization of KCl differ by quality. The latter can be inferred from reading the original papers.

Table 3. Variance components: in - *a priori* information (see text), out - final estimates.

| series code | exp. type | σ_r | | $\sqrt{\gamma_a}$ | | $\sqrt{\gamma_b}$ | |
|-------------|-----------|------------|-------|-------------------|-------|-------------------|-------|
| | | in | out | in | out | in | out |
| RM | P_{tot} | % | 0.035 | % | 2.37 | % | 41.87 |
| WA | P_{tot} | # | 0.032 | # | 0.54 | # | 3.41 |
| FR | P_{tot} | # | 0.032 | # | 0.54 | # | 3.41 |
| D1 | P_{tot} | # | 0.032 | % | 5.97 | # | 3.41 |
| BB1 | P_{tot} | # | 0.032 | # | 0.54 | # | 3.41 |
| PB | P_{tot} | # | 0.032 | # | 0.54 | # | 3.41 |
| NP | P_{tot} | # | 0.032 | # | 0.54 | # | 3.41 |
| SC1 | P_{tot} | # | 0.032 | # | 0.54 | # | 3.41 |
| K | P_{tot} | % | 0.030 | % | 2.02 | % | 30.74 |
| BD | P_{tot} | # | 0.032 | # | 0.54 | % | 14.55 |
| BM | P_{tot} | % | 0.031 | % | 3.11 | % | 75.69 |
| D2 | P_{KE} | # | 0.032 | % | 7.98 | # | 3.41 |
| MW | P_{KE} | # | 0.032 | # | 0.54 | % | 37.14 |
| ZM | P_{KE} | % | 0.092 | % | 0.62 | % | 3.12 |
| BV1 | P_{KE} | # | 0.032 | # | 0.54 | # | 3.41 |
| BV2 | P_{KE} | # | 0.032 | # | 0.54 | # | 3.41 |
| BV3 | P_{KE} | # | 0.032 | # | 0.54 | # | 3.41 |
| MK | P_{KE} | # | 0.032 | % | 12.50 | # | 3.41 |
| NS | P_{KE} | # | 0.032 | # | 0.54 | % | 14.21 |
| ES | P_{KE} | % | 0.022 | % | 4.99 | % | 23.70 |
| KK1 | P_{KE} | % | 0.042 | % | 8.72 | % | 17.47 |
| KK2 | P_{KE} | # | 0.032 | % | 9.94 | # | 3.41 |
| KK3 | P_{KE} | # | 0.032 | % | 8.49 | # | 3.41 |
| vK | P_{KE} | % | 0.160 | % | 0.18 | % | 12.34 |
| N | P_{TR} | % | 0.050 | % | 3.22 | % | 17.39 |
| TW | P_{TR} | % | 0.018 | % | 9.41 | % | 39.21 |
| BB2 | P_{TR} | # | 0.032 | # | 0.54 | # | 3.41 |
| SC2 | P_{TR} | # | 0.032 | # | 0.54 | # | 3.41 |
| MM | P_{TR} | # | 0.032 | # | 0.54 | # | 3.41 |

Table 3 lists the final *a priori* information accepted in Ref. [7] for processing KCl vaporization experimental values as a result of both deviate and expert analyses.

Algorithm for maximizing function (5) is written in terms of $\sigma_{r,i}^2$, $\gamma_{a,i}$, and $\gamma_{b,i}$, and the latter are used in Table 3. Columns `in` show the hypothesis imposed on the maximization of function (5). Symbol # means that these variance components were assumed to be the same in the group of series (for example, $\gamma_{a,i}(\text{WA}) = \gamma_{a,i}(\text{FR}) = \gamma_{a,i}(\text{BB1}) = \dots$), and symbol % implies that this variance component was considered to be an independent unknown. On the whole, there were left 38 unknowns in the variance components besides four unknowns in the physico-chemical model.

Maximization of function (5) under the assumption of Table 3 has given the estimates of the enthalpies and entropies in question (see Table 4, the solution ML) and also that of the variance components (see Table 3, columns `out`).

Function (5) reaches the maximum when the estimates of systematic errors are as small as possible. Under the assumption of the Table 3, it gives that the minimal systematic errors are ascribed to the group of series where the variance component was assumed to be the same. Pragmatically speaking, in our case it meant that when the symbol % was assigned to a variance component this made its final estimate bigger than for that with symbol #. Keeping in mind that the bigger the variance, the smaller the weight, one can say that setting a variance component as an independent unknown (symbol %) was somewhat equivalent to throwing the series out (see series RM, K, BM, *etc.*). Yet, because of two sources for systematic errors this allow us some more flexibility than mere discarding a series. Let us have a look at the series D1, D2, MK, KK2, KK3 (Fig. from 3 to 6). It is possible to say that the slope measured is rather good, and the problem is with the calibration only. The assignment of the different hypotheses to shift and tilt variances for these series (Table 3) permitted us to use the value of the slope and discard the intercept.

Certainly, the results obtained depend on the hypothesis listed in Table 3, and there is a lot of subjectivity in defining that hypothesis. Yet, this is a general situation in mathematical statistics: to obtain results it is necessary to formulate some *a priori* assumptions in addition to the raw experimental values. Also, it should be mentioned that the results obtained in the present paper are rather robust in relation to changes in the hypothesis. As an example, the results of processing the small set of twelve experiments that are in the best accordance with each other (WA, FR, BB1, PB, NP, SC1, BV1, BV2, BV3, BB2, SC2, MM, see Fig. 6) are also given in Table 4 (solution MLs). The analogous variance components were considered to be equal (the total is three unknowns in variance components) in this small set of twelve series. The most differences in solutions ML and MLs are in the standard deviations of the enthalpy and entropy of Reaction (12) that is explained by the smaller total number of the experimental points included in the latter case.

5. COMPARISON WITH THE LEAST SQUARES METHOD

The main point in the present paper is connected not with the maximum likelihood method by itself but with Eq. (2) where two new terms modeling shift and tilt systematic errors has been appeared in addition to the usual reproducibility error. Equation (2) brings forth the block-diagonal structure of the observation dispersion matrix $\mathbf{D}(\varepsilon)$ (Eqs 3 and 4), and here is the main difference with the ordinary and weighted least squares methods (OLS and WLS) that most often are used in practical work. To deal with Eq. (2), it is necessary to employ at least the generalized least squares (GLS).

Equation 7 can be applied to all three methods, OLS, WLS, and GLS, the difference being in the weighted matrix structure only. In its general form as follows

$$SS = \sum_k \sum_l \varepsilon_k \varepsilon_l W_{kl} \quad (18)$$

where indices k and l enumerate the points in all the series, Eq. 7 means the GLS method with the weighted matrix being non-diagonal (a must in the case of the linear error model, Eq. 2). To obtain far more often encountered form for the sum of squares

$$SS = \sum_k \varepsilon_k^2 W_k \quad (19)$$

one has to apply a diagonal weighted matrix that gives the WLS method. And in the OLS

Table 4. The results obtained after processing the KCl vaporization experimental values.

| solution | $\Delta_r H_m^\circ (1) \pm \sigma$ kJ·mol ⁻¹ | | $\Delta_r S_m^\circ (1) \pm \sigma$ J·K ⁻¹ ·mol ⁻¹ | | $\Delta_r H_{dm}^\circ (2) \pm \sigma$ kJ·mol ⁻¹ | | $\Delta_r S_{dm}^\circ (2) \pm \sigma$ J·K ⁻¹ ·mol ⁻¹ | | σ_r | $\sqrt{\gamma_a}$ | $\sqrt{\gamma_b}$ |
|----------|---|-----|---|-----|--|-----|--|-----|------------|-------------------|-------------------|
| | | | | | | | | | | | |
| ML | 206.8 | 0.4 | 133.0 | 0.5 | 37.9 | 2.1 | 24.9 | 1.9 | 0.032 | 0.54 | 3.41 |
| MLs | 207.3 | 0.5 | 133.5 | 0.6 | 33.5 | 2.9 | 21.5 | 2.6 | 0.032 | 0.59 | 3.01 |
| OLSs | 207.2 | 0.3 | 133.4 | 0.3 | 31.6 | 2.0 | 20.0 | 1.6 | 0.036 | 0.00 | 0.00 |

method, all W_k are assumed to be equal to the unit (the weighted matrix is an identity matrix).

Before the minimization of the sum of squares (7), (18), or (19), all the elements of the weighted matrix must be assigned some numerical values, and this sets the limit for the use of the least square method. In the case of the linear error model, Eq. (2), it means that only if the values of $\gamma_{a,i}$ and $\gamma_{b,i}$ are given *a priori* the generalized least squares can be employed. The OLS and WLS methods can be utilized under even more restrictive way, when one *a priori* assigns the zero value to all of the quantities $\gamma_{a,i}$ and $\gamma_{b,i}$.

The maximum likelihood method is one of the means to go forward when there are some unknowns in the weighted matrix. It is worthy of noting that it is possible to put aside all the original hypotheses (see Section 2) that led to Eq. (5) and to consider it as some heuristic empirical tool to determine both parameters (vector Θ) and unknowns in the weighted matrix. From that viewpoint, all modifications of the least square method are the special cases of maximizing function (5) after the application of Eq. (6). Thus, the maximization of (5) can be used whenever the least squares is employed.

Now, let us consider what the consequences in the results may be if one does not account for the systematic errors. To this end, all the examples given in the present paper were also treated by the OLS method. We will start the comparison with the treatment of the simulated data in one-way classification (9) and linear regression (10) discussed in Section 3 (see Fig. 1, 2 and Tables 1, 2).

As discussed above, the application of the OLS method means that the $\gamma_{a,i}$ and $\gamma_{b,i}$ were assigned the zero values *a priori*. In the cases *a*, this happens to be true and the OLS results are completely the same as the ML results. Still, the zeros in the columns $\sqrt{\gamma_a}$ and $\sqrt{\gamma_b}$ for the ML and OLS methods have different meaning. In the ML method the zeros are found as a result of maximizing function (5), and in the OLS method they were set *a priori*.

In cases from *b* to *d*, the application of the OLS method implies that the parameters (vector Θ) are found with the wrong weighted matrix. The pseudo-experimental values are simulated with γ_a and γ_b being non-zero and treated under the assumption that γ_a and γ_b are zeros. Nonetheless, the values of parameters obtained are rather reasonable even in the cases *d* when the level of the introduced systematic errors was pretty high. It is difficult to say though to what extent this is the general case because, in principle, the wrong weighted matrix should mean the wrong parameter estimates. The wrong weighted matrix resulted in the obviously unsatisfactory estimates of the parameter standard deviations. Here, the results obtained by the OLS method are far from the true ones and those obtained by the ML method.

The variance estimates obtained by the ML method are downward biased [1, 2]. Probably, it can be seen from Tables 1 and 2 where most of the ML's γ_a and γ_b estimates are lower than the true ones although the sample is too small for definite conclusions. Still, even downward biased estimated of γ_a and γ_b obtained in the ML method are better than the assignment of the zero values to them *a priori* in the OLS method.

The results displayed in Table 1 and 2 suggest us a rule of thumb as follows. If there are systematic errors, *i.e.* the between-errors are greater than the within-errors, one may expect that the OLS treatment probably gives not bad estimates of parameters but with underestimated variances (the derived confidence limits are too optimistic).

Let us now discuss the application of the OLS method to the vaporization of KCl (see Fig. from 3 to 6 and Table 4). One of the anonymous referees recommended the naive algorithm: 'Perform a least squares fit to all points, compute apparent tilt and shift errors for all experiments, reject the poorest experiment(s), and iterate this screening. Perform the final fit with those data which remain and which thus consistent.' Unfortunately, in the case in question this is not possible. The first action is quite evident - there are twelve experiments (WA, FR, BB1, PB, NP, SC1, BV1, BV2, BV3, BB2, SC2, MM), the central spot on Fig. 6, which are in the best accordance with each other. The problem is that even for these twelve experiments the between-errors are greater than the within-errors. The latter is clearly indicated by the MLs solution performed for this small set of the series (see Table 4). In principle, it is possible to go further but in practice it would mean that there are several self-consistent sets each comprising two-four series. Also, the expert opinion tells us that these twelve experiments possess the same quality.

Under these circumstances, the OLS method has been applied to that small set of twelve experiments (the solution OLSs in Table 4). Having compared the OLSs results obtained with the MLs solution, it is possible to say that in the KCl case the rule of thumb works -- the enthalpies and entropies obtained are rather good but their standard deviations are too small.

No one of the mathematical statistics methods can be applied to raw experimental data without using some expert information. In the least squares method, an expert must express his/her knowledge as quantitative information, *i.e.* to assign some numerical values to all the elements of the weighted matrix. This changes in the ML method because it is already possible to leave unknowns in the weighted matrix. The ML method allows the expert to give rise to the qualitative information only (see Table 3 and Section 4), and thus, to remove a burden of yielding quantitative information from the expert. In the framework of the ML method, the expert have to separate the series into the groups with the same quality, and after that, he/she can leave the determination of the weights to the numerical procedure.

Certainly, the ML method takes more processor time than the least squares. In the present work to obtain ML estimates, the time required was from four to ten times more depending on the choice of the initial guesses than for the least squares estimates. Still, the calculation time to process the KCl vaporization data was just a few minutes on my 386DX40 IBM PC (by modern standards, such a computer is considered as rather slow). Thus, one can hardly name the ML method in the case of the linear error model (2) as computationally intensive.

6. ALGORITHM

The main problem in finding the maximum of (5) is to compute the inverse dispersion matrix $\mathbf{D}(\boldsymbol{\varepsilon})^{-1}$. However, in the case of the linear error model (2) it happens to be possible to obtain the inverse matrix in the closed form. This allows the simple but effective algorithm to obtain.

6.1. Some expressions for the dispersion matrix in the closed form

Because of the block-diagonal structure of $\mathbf{D}(\boldsymbol{\varepsilon})$ (Eqs 3 and 4), one can write

$$\mathbf{D}(\boldsymbol{\varepsilon})^{-1} = \text{diag}\{\mathbf{V}_j^{-1}\}$$

$$SS = \sum_i \boldsymbol{\varepsilon}_i' \mathbf{V}_i^{-1} \boldsymbol{\varepsilon}_i$$

$$\ln \{\det \mathbf{D}(\boldsymbol{\varepsilon})\} = \sum_i \ln \{\det \mathbf{V}_i\}$$

After change of $\sigma_{a,i}^2$ and $\sigma_{b,i}^2$ to $\gamma_{a,i}$ and $\gamma_{b,i}$, Eq. (4) takes the form

$$\mathbf{V}_i = \sigma_{r,i}^2 (\mathbf{1}_i + \gamma_{a,i} \mathbf{1}_i \mathbf{1}_i' + \gamma_{b,i} \mathbf{x}_i \mathbf{x}_i')$$

and in this notation the following is held

$$\mathbf{V}_i^{-1} = \sigma_{r,i}^{-2} \{\mathbf{I}_i - \gamma_{a,i} (1 + N_i \gamma_{a,i})^{-1} \mathbf{1}_i \mathbf{1}_i' - \gamma_{b,i} (1 + P_i \gamma_{b,i})^{-1} \mathbf{x}_i \mathbf{x}_i'\}$$

$$\det \mathbf{V}_i = (\sigma_{r,i}^2)^{N_i} (1 + N_i \gamma_{a,i})(1 + P_i \gamma_{b,i})$$

This can be tested directly. Note that $\mathbf{1}_i' \mathbf{1}_i = N_i$, $\mathbf{x}_i' \mathbf{x}_i = P_i$ and $\mathbf{1}_i' \mathbf{x}_i = 0$.

Thus, the function L becomes

$$L = - \sum_i \{N_i \ln \sigma_{r,i}^2 + \ln(1 + N_i \gamma_{a,i}) + \ln(1 + P_i \gamma_{b,i})\} - SS \quad (20)$$

where SS is the generalized sum of squared deviations, which, in turn, is expressed as

$$SS = \sum_i \{ \sum_j \varepsilon_{ij}^2 - \gamma_{a,i} (1 + N_i \gamma_{a,i})^{-1} (\sum_j \varepsilon_{ij})^2 - \gamma_{b,i} (1 + P_i \gamma_{b,i})^{-1} [\sum_j \varepsilon_{ij} (x_{ij} - \bar{x}_i)]^2 \} / \sigma_{r,i}^2 \quad (21)$$

It is possible to obtain the alternative expression for SS with the use of the matrix $\mathbf{D}(\boldsymbol{\varepsilon})^{-1/2}$ such that

$$\mathbf{D}(\boldsymbol{\varepsilon})^{-1/2} \mathbf{D}(\boldsymbol{\varepsilon})^{-1/2} = \mathbf{D}(\boldsymbol{\varepsilon})^{-1}$$

It has also a block-diagonal form

$$\mathbf{D}(\boldsymbol{\varepsilon})^{-1/2} = \text{diag} \{ \mathbf{V}_i^{-1/2} \}$$

and one can check directly that

$$\mathbf{V}_i^{-1/2} = \{ \mathbf{I}_i - \delta_{a,i} \mathbf{1}_i \mathbf{1}_i' - \delta_{b,i} \mathbf{x}_i \mathbf{x}_i' \} / \sigma_{r,i}$$

where

$$\delta_{a,i} = \{ 1 - (1 + N_i \gamma_{a,i})^{-1/2} \} / N_i$$

$$\delta_{b,i} = \{ 1 - (1 + P_i \gamma_{b,i})^{-1/2} \} / P_i$$

Then, the following is held

$$\begin{aligned} SS &= \sum_i (\mathbf{V}_i^{-1/2} \boldsymbol{\varepsilon}_i)' (\mathbf{V}_i^{-1/2} \boldsymbol{\varepsilon}_i) = \\ &= \sum_i \sum_j \{ [\varepsilon_{ij} - \delta_{a,i} (\sum_j \varepsilon_{ij}) - \delta_{b,i} (x_{ij} - \bar{x}_i) (\sum_j \varepsilon_{ij} (x_{ij} - \bar{x}_i))] / \sigma_{r,i} \}^2 \end{aligned} \quad (22)$$

Such a form is analogous to the ordinary sum of squares (compare with Eq. 19) and permits us to apply well-known non-linear algorithms (*e.g.* Levenberg-Marquardt algorithm) to the minimization of the generalized sum of squares SS over unknown parameters.

6.2. Finding the maximum of the likelihood function

Equation (22) permits us to apply somewhat "iterative" least squares method to maximizing function (5).

It is suggested to start with some *a priori* estimates for variance components. In my work, the initial guesses of variance components, $\sigma_{r,i}^2 = 1$, $\gamma_{a,i} = 0$ and $\gamma_{b,i} = 0$ are proved to be good. Then, the first step is to find the maximum of (5) with all the variance components kept constant (the matrix $\mathbf{D}(\boldsymbol{\varepsilon})$ is given). This coincides with the minimization of SS over unknown parameters of the physico-chemical model (the latter are hidden inside of $\boldsymbol{\varepsilon}_{ij}$), *i.e.* with the generalized least squares method (Eq. 7). To this end, Eq. (22) make it possible to employ well-developed algorithms for the non-linear minimization of the ordinary sum of squared deviations. In my work, the subroutine ZXSSQ from the IMSL library, which realizes the finite difference Levenberg-Marquardt algorithm was utilized.

After that, it is proposed to keep constant the values of $\boldsymbol{\varepsilon}_{ij}$ obtained in the previous step and to find the maximum of Eq. (5) over the variance components $\sigma_{r,i}^2$, $\gamma_{a,i}$ and $\gamma_{b,i}$ when the values of $\sigma_{r,i}^2$ are assumed to be positive and these of $\gamma_{a,i}$ and $\gamma_{b,i}$ are non-negative. It can be done by equating to zero the derivatives of L with respect to the variance components (their form will be simpler if the sum of squares given by Eq. 21 and not Eq. 22 is used) and solving the non-linear equations obtained by simple iterations (see Ref. [2] for details). If a value of $\gamma_{a,i}$ or $\gamma_{b,i}$ gets negative it is changed to zero.

Now, the whole procedure can be iterated with the variance component estimates obtained in the second step as the initial guesses for the first step until convergence is achieved. Unfortunately, I have no proof that this algorithm converges in all cases. All I can say that it worked in my applications.

For the models (9) and (10), the algorithm is programmed in the utility LINEAR [6] and you are welcome to try it.

7. ACKNOWLEDGMENT

The work reported in this paper was supported by Grant No. 93-03-5650 of the Russian Foundation for fundamental science. I also would like to thank two anonymous referees for their useful and helpful comments.

8. REFERENCES

- [1] Rao, C.R.; Kleffe, J. "Estimation of variance components and applications," North-Holland, Amsterdam, (North-Holland Series in Statistics and probability, v.3), 1988, 370 pp.
- [2] Harville, D.A. "Maximum-likelihood approaches to variance component estimation and to related problems," J. Amer. Statist. Assoc., 72, 320-340 (1977).

- [3] Markova, E.V.; Denisov, V.I.; Poletaeva, I.A.; Ponomarev, V.V. "Variance analysis and plan synthesis with a computer," (in Russian), Moscow, Nauka, 1982, 195 pp.
- [4] Paule, R.C.; Mandel, J. "Consensus Values, Regressions, and Weighting Factors," J. Res. NIST, 94, 197-203 (1989).
- [5] Vonesh, E.F.; Carter, R.L. "Mixed-effects Nonlinear Regression for Unbalanced Repeated Measures," Biometrics, 48, 1-18 (1992).
- [6] Rudnyi, E.B. "GENERATE and LINEAR - MS-DOS utilities for processing of experimental series with systematic errors," 1994, available at <http://www.chem.msu.su/~rudnyi/pcmler.zip>
- [7] Rudnyi, E.B.; Bonnell, D.W.; Hastie D.W. "Vaporization thermodynamics of KCl. Combined processing of total vapor pressure and evaporation rates," (in Russian), Vestn. Mosk. Univ., ser. Khimiya, 35, 291-308 (1994). English transl. in Moscow Univ. Chem. Bul.