Acta Univ. Palacki. Olomuc., Fac. rer. nat.,
Mathematica **51**, 1 (2012) 5–27

Optimization of parameters in the Menzerath–Altmann law*

Jan ANDRES, Lubomír KUBÁČEK,
Jitka MACHALOVÁ, Michaela TUČKOVÁ

[1] *Department of Mathematical Analysis and Applications of Mathematics*
Faculty of Science, Palacký University
17. listopadu 12, 771 46 Olomouc, Czech Republic
e-mail: jan.andres@upol.cz, lubomir.kubacek@upol.cz,
jitka.machalova@upol.cz, michaela.tuckova@upol.cz

(Received January 17, 2012)

Abstract

Four formulas of the Menzerath–Altmann law are tested from the point of view of their applicability and suitability. The accuracy of related approximations of measured data is examined by the least square method at first. Then the accuracy of calculated parameters in the formulas under consideration is compared statistically. The influence of neglecting parameter c is investigated as well. Finally, the obtained results are discussed by means of an illustrative example from quantitative linguistics.

Key words: Menzerath–Altmann's law, accuracy of data approximations, accuracy of parameter estimates, mean square error, least square method, optimization, comparison

2000 Mathematics Subject Classification: 33F05, 65D07, 65D15, 91F20

1 Introduction

The Menzerath–Altmann law (shortly, MAL) is with no doubts one of the milestones of quantitative linguistics. Its verbal (heuristic) form says that "the longer a language construct is, the shorter its components (constituents) are." By a construct we mean a unit on a higher language level (e.g. a *clause*, i.e. an autonomous unit in terms of its pragmatics, semantic construction and grammar which is based on a finite verb form) and by a constituent we mean a unit

*Supported by the Council of the Czech Government (MSM 6198959214).

on a lower language level (e.g. a word). The length x of a construct is measured in its constituents, while y is the average length of its constituents measured in units on the nearest lower language level. For instance, x can be a length of a clause measured in the number of words and y then denotes the average number of composing words measured in the number of syllables.

Although the length x of a construct is (unlike $y \in (0,\infty)$) always a positive integer, i.e. $x \in \mathbb{N}$, for the sake of derivation of the mathematical "continuous" form of MAL, we will nevertheless consider x to be also positively real-valued, i.e. $x \in (0,\infty)$.

Hence, recalling the mathematical derivation of MAL due to G. Altmann [1], [2], [20], the relative rate of the change of y can be expressed as $\frac{\dot{y}}{y}$, where $\dot{y} = \frac{dy}{dx}$ is obviously the derivative of y w.r.t. x. According to the verbal formulation above, this rate is inversely proportional to x up to an additive constant $c \in \mathbb{R}$, i.e.

$$\frac{\dot{y}}{y} = \frac{-b}{x} + c, \tag{1}$$

where $-b$ is the proportionality coefficient. Let us note that it is convenient to take it with the minus sign, as we shall see later.

Integrating the ordinary differential equation (1) w.r.t. x, we obtain the equation

$$\ln y = -b \ln x + cx + C,$$

where $C \in \mathbb{R}$ is another additive constant, i.e. after delogarithmization, we get

$$y = e^C x^{-b} e^{cx}.$$

Thus, the general solution of (1) takes the form

$$y(x) = Ax^{-b}e^{cx}, \quad \text{where } A = e^C > 0. \tag{2}$$

This is so called *complete formula* of MAL.

Let us note that in many empirical studies concerning sentence or clause structures, and all the better in supra-sentence structures like semantic constructs, only its hyperbolic part is used, namely ($c = 0$)

$$y = Ax^{-b}, \tag{3}$$

see e.g. [9], [10]. Here, we therefore speak about the *truncated formula* of MAL.

On the other hand, it was demonstrated that the role of the exponential parts, which can be omitted in the case of semiotically higher levels, increases with a decreasing linguistic level, and so it is usually not omitted (for instance, in the case of words and syllables).

In order to simplify the computation of parameters b, c in (2) or b in (3), we can still put $A := \frac{y(1)}{e^c} = \frac{y_1}{e^c}$ in (2), i.e.

$$y = y_1 x^{-b} e^{c(x-1)}, \tag{4}$$

or $A := y(1) = y_1$ in (3), i.e.

$$y = y_1 x^{-b}. \tag{5}$$

The main aim of the present paper is to show both theoretically as well as practically (by means of an illustrative linguistic example) how much the formulas (2)–(5) differ each from other. For the first glance, it might be expected that the optimal results must be connected with the complete formula (2), while the roughest are those related to the truncated formula (5). However, as we shall see, the situation will not be so simple but rather delicate for such a statement. It will become more transparent and maybe also a bit surprising by means of the least square method applied below for numerical computation of parameters A, b, c in (2)–(5) in Section 2 and their statistical verification in Sections 3 and 4.

For more details concerning various aspects of MAL, we recommend e.g. the papers [1], [2], [3], [4], [6], [8]–[10], [11]–[13], [15], [16], [18], [20], [21]. In [20], the whole panorama of further variants of linguistic formulas of this sort is presented.

2 Accuracy of approximations of measured data

Let us consider separately the following cases:

I) $y = y_1 x^{-b}$, i.e. $A = y(1) = y_1$ and $c = 0$, (cf. (5)),

II) $y = Ax^{-b}$, i.e. $c = 0$, (cf. (3)),

III) $y = y_1 x^{-b} e^{c(x-1)}$, i.e. $Ae^c = y(1) = y_1$, (cf. (4)),

IV) $y = Ax^{-b} e^{cx}$, (cf. (2)).

ad I) (one free parameter b)

Logarithmizing the equation (5), we obtain

$$\ln y = \ln y_1 - b \ln x,$$

i.e.

$$\ln y_i = \ln y_1 - b \ln x_i, \quad i = 1, 2, \ldots,$$

where $x_i = i$ denotes the length of the i-th construct and y_i denotes the length of the i-th constituent of x_i.

Hence, denoting still $u_i := \ln x_i$ and $v_i := \ln y_i$, we can minimize the function

$$\psi_I(b) := \sum_i w_i \left[v_1 - bu_i - v_i\right]^2$$

w.r.t. b, where $w_i = \frac{z_i}{\sum_i z_i}$ are the weights corresponding to the i-th relative frequency $\frac{z_i}{\sum_i z_i}$.

Putting

$$\dot{\psi}_I(b) = 2 \sum_i w_i \left[v_1 - bu_i - v_i\right](-u_i) = 0,$$

we arrive at

$$b_I = \frac{v_1 \sum_i w_i u_i - \sum_i w_i u_i v_i}{\sum_i w_i u_i^2} = \frac{\ln y_1 \sum_i w_i \ln x_i - \sum_i w_i \ln x_i \ln y_i}{\sum_i w_i \left(\ln x_i\right)^2}. \quad (6)$$

Since

$$\ddot{\psi}_I(b) = 2\sum_i w_i u_i^2 > 0,$$

we really have that

$$\min_b \psi_I(b) = \psi_I(b_I).$$

Moreover, the related least square value Δ_I takes the form

$$\Delta_I := \sum_i w_i \left[y(x_i) - y_i\right]^2 = \sum_i w_i \left[y_1 x_i^{-b_I} - y_i\right]^2. \tag{7}$$

Remark 1 Let us note that b_I can slightly differ from $\widehat{b_I}$ satisfying

$$\min_b \sum_i w_i \left[y_1 x_i^{-b} - y_i\right]^2 = \sum_i w_i \left[y_1 x_i^{-\widehat{b_I}} - y_i\right]^2,$$

but statistically b_I, and subsequently also Δ_I, can be acceptable (see e.g. [14, pp. 219–225]).

Example 1 Considering the Czech translation by Slavík of Poe's Raven, the following data were calculated in [5]:

x_i	y_i	z_i
1	3.4444	9
2	1.8548	31
3	1.5256	26
4	1.8068	22
5	1.9467	15
6	1.8810	7
7	1.9091	11
8	1.8250	5
9	1.9899	11
10	1.6000	3
11	1.7955	4
12	1.8750	4

x_i	y_i	z_i
13	2.0769	1
14	1.7500	2
15	2.2000	1
16	1.6875	1
17	1.6471	1
18	1.8333	1
19		0
20		0
21		0
22		0
23	1.8696	1
24	1.8333	1

Here, x_i indicates that the clause consists of i words (i.e. that the length of i-th sentence is i), its frequence in the given translation is z_i and y_i denotes the average length of composing words of such sentences.

Applying formulas (6) and (7), b_I and $\Delta_I = \Delta_I(b_I)$ can be easily calculated as follows:

$$b_I \doteq 0.3574, \qquad \Delta_I \doteq 0.3014.$$

Because of $A_I = y_1$, we trivially have $A_I \doteq 3.4444$.

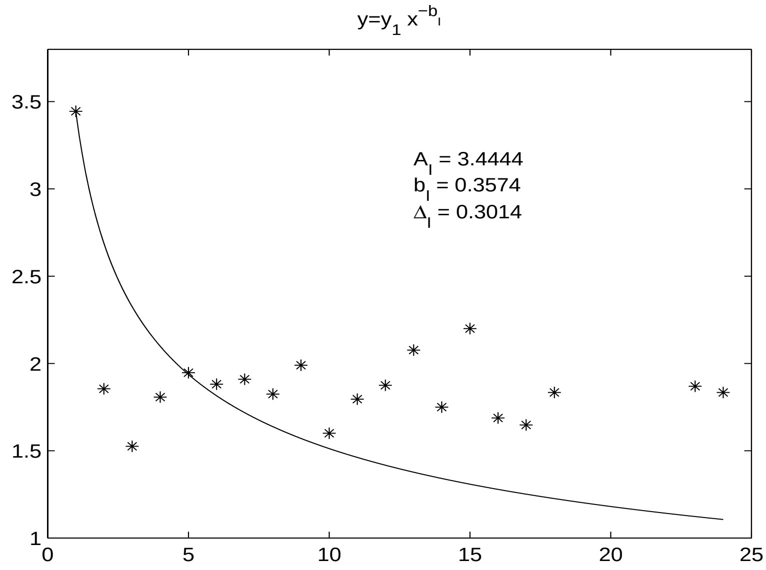

Figure 1: Calculated data y_i (see the tables above) versus the values $y(x_i) = 3.4444i^{-0.3574}$, $i = 1, 2, \ldots, 24$.

ad II) (two free parameters A, b)

Logarithmizing the equation (3), we obtain

$$\ln y = \ln A - b \ln x,$$

i.e.

$$\ln y_i = \ln A - b \ln x_i, \quad i = 1, 2, \ldots,$$

where the symbols x_i and y_i have the same meaning as above.

Hence, denoting again $u_i := \ln x_i$, $v_i := \ln y_i$ and newly $a := \ln A$, we can minimize the function

$$\psi_{II}(a, b) := \sum_i w_i \left[a - bu_i - v_i\right]^2$$

w.r.t. $a = \ln A$, b, where the weights $w_i = \frac{z_i}{\sum_i z_i}$ are the same as in the case I).

Putting

$$\frac{\partial \psi_{II}}{\partial a} = 2 \sum_i w_i \left[a - bu_i - v_i\right] = 0,$$

$$\frac{\partial \psi_{II}}{\partial b} = 2 \sum_i w_i \left[a - bu_i - v_i\right](-u_i) = 0,$$

we arrive at

$$A_{II} = \exp\left(\frac{\sum_i w_i \ln x_i \sum_i w_i (\ln x_i)(\ln y_i) - \sum_i w_i (\ln x_i)^2 \sum_i w_i \ln y_i}{\left[\sum_i w_i \ln x_i\right]^2 - \sum_i w_i (\ln x_i)^2}\right), \quad (8)$$

$$b_{II} = \frac{\sum_i w_i (\ln x_i)(\ln y_i) - \sum_i w_i \ln x_i \sum_i w_i \ln y_i}{\left[\sum_i w_i \ln x_i\right]^2 - \sum_i w_i (\ln x_i)^2}. \quad (9)$$

Since

$$\frac{\partial^2 \psi_{II}}{\partial a^2} = 2\sum_i w_i = 2 > 0$$

and (in view of the Schwarz inequality)

$$\frac{\partial^2 \psi_{II}}{\partial a^2}\frac{\partial^2 \psi_{II}}{\partial b^2} - \left[\frac{\partial^2 \psi_{II}}{\partial a \partial b}\right]^2 = 4\sum_i w_i u_i^2 - 4\left[\sum_i w_i u_i\right]^2 > 0,$$

we really have that

$$\min_{a,b} \psi_{II}(a,b) = \psi_{II}\left(\ln A_{II}, b_{II}\right).$$

Moreover, the related least square value Δ_{II} takes the form

$$\Delta_{II} := \sum_i w_i \left[A_{II} x_i^{-b_{II}} - y_i\right]^2. \quad (10)$$

Example 2 Consider the same tables as in Example 1 and let us note that the analogy of Remark 1 holds here as well.

Hence, applying formulas (8), (9) and (10), the values of A_{II}, b_{II} and Δ_{II} can be calculated as follows:

$$A_{II} \doteq 2.0815, \quad b_{II} \doteq 0.0753, \quad \Delta_{II} \doteq 0.1453.$$

ad III) (two free parameters b, c)

Logarithmizing the equation (4), we obtain

$$\ln y = \ln y_1 - b \ln x + c\,(x-1)\,,$$

i.e.

$$\ln y_i = \ln y_1 - b \ln x_i + c\,(x_i - 1)\,, \quad i = 1, 2, \ldots,$$

where the symbols x_i and y_i have the same meaning as above.

Hence, denoting again $u_i := \ln x_i$ ($\Rightarrow x_i = \exp(u_i)$) and $v_i = \ln y_i$, we can minimize the function

$$\psi_{III}(a,b) := \sum_i w_i \left[v_1 - b u_i + c\,(x_i - 1) - v_i\right]^2$$

w.r.t. b, c, where the weights $w_i = \frac{z_i}{\sum_i z_i}$ are the same as above.

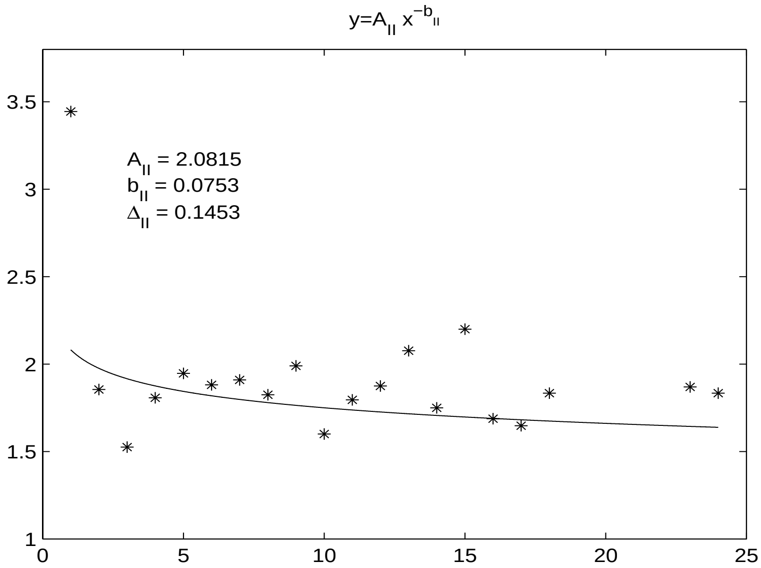

Figure 2: Calculated data y_i (see the tables above) versus the values $y(x_i) = 2.0815 i^{-0.0753}$, $i = 1, 2, \ldots, 24$.

Putting

$$\frac{\partial \psi_{III}}{\partial b} = 2 \sum_i w_i \left[v_1 - b u_i + c\,(x_i - 1) - v_i \right] (-u_i) = 0,$$

$$\frac{\partial \psi_{III}}{\partial c} = 2 \sum_i w_i \left[v_1 - b u_i + c\,(x_i - 1) - v_i \right] (x_i - 1) = 0,$$

we arrive at ($A_{III} = \frac{y_1}{e^{c_{III}}}$)

$$b_{III} = \frac{\sum_i w_i (x_i - 1)^2 \sum_i w_i \ln x_i (\ln y_1 - \ln y_i)}{\sum_i w_i (\ln x_i)^2 \sum_i w_i (x_i - 1)^2 - \left[\sum_i w_i (x_i - 1) \ln x_i\right]^2} + \frac{\sum_i w_i (x_i - 1) \ln x_i \sum_i w_i (x_i - 1)(\ln y_i - \ln y_1)}{\sum_i w_i (\ln x_i)^2 \sum_i w_i (x_i - 1)^2 - \left[\sum_i w_i (x_i - 1) \ln x_i\right]^2}, \tag{11}$$

$$c_{III} = \frac{\sum_i w_i (x_i - 1) \ln x_i \sum_i w_i \ln x_i (\ln y_1 - \ln y_i)}{\sum_i w_i (\ln x_i)^2 \sum_i w_i (x_i - 1)^2 - \left[\sum_i w_i (x_i - 1) \ln x_i\right]^2} + \frac{\sum_i w_i (\ln x_i)^2 \sum_i w_i (x_i - 1)(\ln y_i - \ln y_1)}{\sum_i w_i (\ln x_i)^2 \sum_i w_i (x_i - 1)^2 - \left[\sum_i w_i (x_i - 1) \ln x_i\right]^2}. \tag{12}$$

Since

$$\frac{\partial^2 \psi_{III}}{\partial b^2} = 2 \sum_i w_i u_i^2 = 2 \sum_i w_i (\ln x_i)^2 > 0$$

and (in view of the Schwarz inequality)

$$\frac{\partial^2\psi_{III}}{\partial b^2}\frac{\partial^2\psi_{III}}{\partial c^2} - \left[\frac{\partial^2\psi_{III}}{\partial b\partial c}\right]^2 =$$

$$= 4\sum_i w_i\,(\ln x_i)^2 \sum_i w_i\,(x_i-1)^2 - 4\left[\sum_i w_i\,(x_i-1)\ln x_i\right]^2 > 0,$$

we really have that

$$\min_{b,c}\psi_{III}(b,c) = \psi_{III}\,(b_{III}, c_{III})\,.$$

Moreover, the related least square value Δ_{III} takes the form

$$\Delta_{III} := \sum_i w_i\left[y_1 x_i^{-b_{III}} e^{c_{III}(x_i-1)} - y_i\right]^2. \qquad (13)$$

Example 3 Consider the same tables as in Example 1 and let us note that the analogy of Remark 1 holds here as well.

Hence, applying formulas (11), (12) and (13), the values of b_{III}, c_{III} and Δ_{III} can be calculated as follows:

$$A_{III} \doteq 3.1424, \quad b_{III} \doteq 0.6788, \quad c_{III} \doteq 0.0918, \quad \Delta_{III} \doteq 0.1399.$$

ad IV) (three free parameters A, b, c)

Logarithmizing the equation (2), we obtain

$$\ln y = \ln A - b\ln x + cx,$$

i.e.

$$\ln y_i = \ln A - b\ln x_i + cx_i, \quad i = 1, 2, \ldots,$$

where the symbols x_i and y_i have the same meaning as above.

Hence, denoting again $u_i := \ln x_i$ ($\Rightarrow x_i = \exp(u_i)$), $v_i := \ln y_i$ and $a := \ln A$, we can minimize the function

$$\psi_{IV}(a,b,c) := \sum_i w_i\,[a - bu_i + cx_i - v_i]^2$$

w.r.t. a, b, c, where the weights $w_i = \frac{z_i}{\sum_i z_i}$ are the same as above.

Putting

$$\frac{\partial\psi_{IV}}{\partial a} = 2\sum_i w_i\,[a - bu_i + cx_i - v_i] = 0,$$

$$\frac{\partial\psi_{IV}}{\partial b} = 2\sum_i w_i\,[a - bu_i + cx_i - v_i]\,(-u_i) = 0,$$

$$\frac{\partial\psi_{IV}}{\partial c} = 2\sum_i w_i\,[a - bu_i + cx_i - v_i]\,x_i = 0,$$

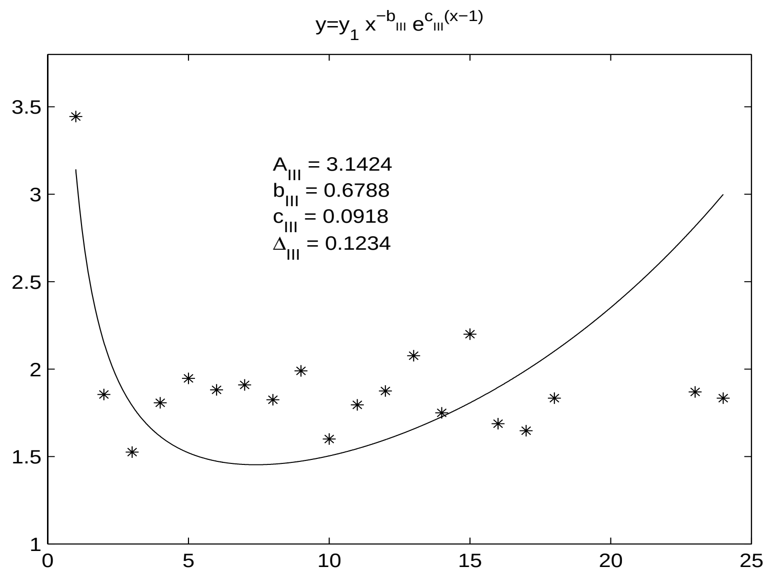

Figure 3: Calculated data y_i (see the tables above) versus the values $y(x_i) = 3.1424 i^{-0.6788} e^{0.0918(i-1)}$, $i = 1, 2, \ldots, 24$. Since the related function is not decreasing, it becomes linguistically uninterpretable.

we arrive at ($A_{IV} = \exp(a_{IV})$)

$$\begin{aligned}
a_{IV} = \frac{1}{\tau}\Bigg\{&\sum_i w_i \ln y_i \left[\left(\sum_i w_i x_i \ln x_i\right)^2 - \sum_i w_i x_i^2 \sum_i w_i (\ln x_i)^2\right] \\
+ &\sum_i w_i \ln x_i \ln y_i \left[\sum_i w_i x_i^2 \sum_i w_i \ln x_i - \sum_i w_i x_i \sum_i w_i x_i \ln x_i\right] \\
+ &\sum_i w_i x_i \ln y_i \left[\sum_i w_i x_i \sum_i w_i (\ln x_i)^2 - \sum_i w_i \ln x_i \sum_i w_i x_i \ln x_i\right]\Bigg\}, \qquad (14)
\end{aligned}$$

$$\begin{aligned}
b_{IV} = \frac{1}{\tau}\Bigg\{&\sum_i w_i \ln y_i \left[\sum_i w_i x_i \sum_i w_i x_i \ln x_i - \sum_i w_i x_i^2 \sum_i w_i \ln x_i\right] \\
&+ \sum_i w_i \ln x_i \ln y_i \left[\sum_i w_i x_i^2 - \left(\sum_i w_i x_i\right)^2\right] \\
&+ \sum_i w_i x_i \ln y_i \left[\sum_i w_i x_i \sum_i w_i \ln x_i - \sum_i w_i x_i \ln x_i\right]\Bigg\}, \qquad (15)
\end{aligned}$$

$$
\begin{aligned}
c_{IV} = \frac{1}{\tau}\Bigg\{ & \sum_i w_i \ln y_i \left[\sum_i w_i x_i \sum_i w_i (\ln x_i)^2 - \sum_i w_i \ln x_i \sum_i w_i x_i \ln x_i\right] \\
& + \sum_i w_i \ln x_i \ln y_i \left[\sum_i w_i x_i \ln x_i - \sum_i w_i x_i \sum_i w_i \ln x_i\right] \\
& + \sum_i w_i x_i \ln y_i \left[\left(\sum_i w_i \ln x_i\right)^2 - \sum_i w_i (\ln x_i)^2\right]\Bigg\}, \qquad (16)
\end{aligned}
$$

where

$$
\begin{aligned}
\tau = & \left(\sum_i w_i x_i \ln x_i\right)^2 + \left(\sum_i w_i \ln x_i\right)^2 \sum_i w_i x_i^2 + \left(\sum_i w_i x_i\right)^2 \sum_i w_i (\ln x_i)^2 \\
& - \sum_i w_i (\ln x_i)^2 \sum_i w_i x_i^2 - 2\sum_i w_i \ln x_i \sum_i w_i x_i \sum_i w_i x_i \ln x_i .
\end{aligned}
$$

We can namely prove again (see e.g. [14, p. 83]) that

$$
\min_{a,b,c} \psi_{IV}(a,b,c) = \psi_{IV}\left(a_{IV}, b_{IV}, c_{IV}\right).
$$

Moreover, the related the least square value takes the form

$$
\Delta_{IV} := \sum_i w_i \left[A_{IV} x_i^{-b_{IV}} e^{c_{IV} x_i} - y_i\right]^2 . \qquad (17)
$$

Example 4 Consider the same tables as in Example 1 and let us note that the analogy of Remark 1 holds here as well.

Hence, applying formulas (15), (16) and (17), the values of A_{IV}, b_{IV}, c_{IV} and Δ_{IV} can be calculated as follows:

$$
A_{IV} \doteq 2.2939, \quad b_{IV} \doteq 0.3094, \quad c_{IV} \doteq 0.0442, \quad \Delta_{IV} \doteq 0.1085.
$$

Summing up the above calculations, we can give the following theorem.

Theorem 1 *Approximative values of parameters A, b, c in formulas (2)–(5) of MAL can be obtained from the lengths of associated constructs $x_1, x_2, \ldots,$ their frequences $z_1, z_2, \ldots,$ and constituents $y_1, y_2, \ldots,$ by means of a regression method in the forms (6) and (8), (9) and (11), (12) and (14)–(16), respectively. The appropriate least square values take the forms (7) and (10) and (13) and (17), respectively.*

3 Comparison of accuracy of parameter estimations

In order to compare formulas (2)–(5), the *experiment design theory* should be employed (see e.g. [17]). This technique is closely related to the previous one but

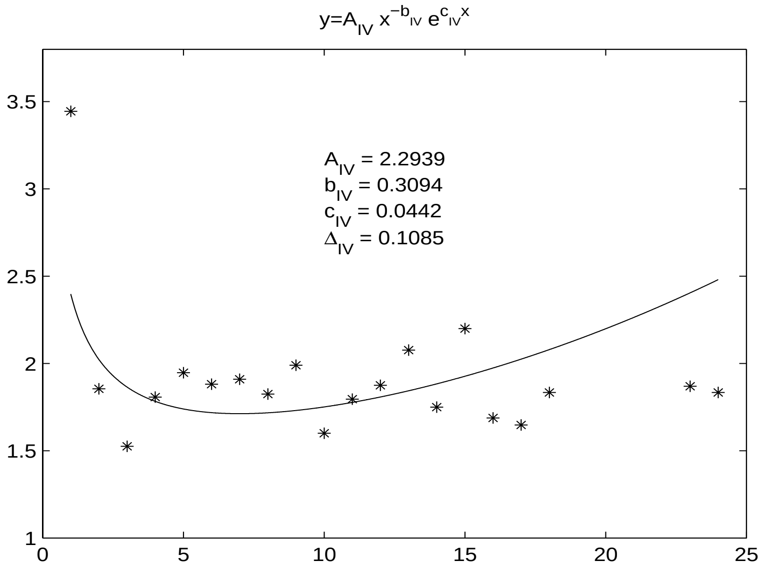

Figure 4: Calculated data y_i (see the tables above) versus the values $y(x_i) = 2.2939i^{-0.3094}e^{0.0442i}$, $i = 1, 2, \ldots, 24$. Since the related function is not decreasing, it becomes linguistically uninterpretable.

the output is different. It does not concern the information about conformity of data with the applied approximation formula, but the value of variance of the free parameters estimates (in statistical terms, we speak about unknown parameters) $\widehat{\mathbf{\Theta}}_I := b$ or $\widehat{\mathbf{\Theta}}_{II} := (A, b)'$ or $\widehat{\mathbf{\Theta}}_{III} := (b, c)'$ or $\widehat{\mathbf{\Theta}}_{IV} := (A, b, c)'$ which appear in the respective formulas (2)–(5). The prime symbol denotes as usually the transposition. Althought this analysis is again based on the least square method, it is necessary to introduce adequate terms because of input variables constructed in a different way. After logarithmizing the equations (2)–(5), we obtain four linear regression models. Their general form is

$$y \sim_n \left(\mathbf{F\Theta}, \sigma^2\mathbf{I}\right), \tag{18}$$

where y has the same meaning as above, n is the number of measurement points, $\mathbf{F}$ is a *design matrix* given by the concrete formula (2) or (3) or (4) or (5), $\mathbf{\Theta}$ is an appropriate *vector of unknown parameters*, σ^2 is a *measurement error* and $\mathbf{I}$ denotes the *unit matrix.* We assume the independence of random variables y_i, $i = 1, \ldots, n$.

Without any loss of generality, we can now restrict ourselves to a specific area of the experiment realization which is known as the *set of experimental points*

$$x = \{x_1, \ldots, x_n\}.$$

The symbol $x_i = i$ has the same meaning as above and in order to guarantee the sufficient richness of the experimental set, we consider the total number of measurements n to be equal to 10 000.

The following analysis is based on the validity of the formula (see [14, p. 50])

$$\mathrm{Var}(\widehat{\mathbf{\Theta}}(\mathbf{Y}_\delta), N) := \frac{\sigma^2}{N}\mathbf{M}^{-1}(\delta) \tag{19}$$

whose proof can be found in ([19]). The symbol N denotes the total number of measurements at the points x_i, i.e. $N = \sum_i z_i$. It says that the covariance matrix estimate of unknown parameters (contained in a regression model) is equal to the product of the expression $\frac{\sigma^2}{N}$ and the inverted matrix $\mathbf{M}(\delta)$, called *information matrix*, namely

$$\mathbf{M}(\delta) := \sum_{i=1}^{n} \delta_i w_i \mathbf{f}_i \mathbf{f}_i'. \tag{20}$$

The form (20) of an information matrix is given by several variables. In particular, the *design of experiment* is a function

$$\delta_i : x_i \to [0, 1], \quad \text{where } \sum_{i=1}^{n} \delta_i = 1, \ i = 1, \ldots, n,$$

which determines a significance of the experiment at various points of the experimental set. In our case, the significance is the same for all experimental points. Therefore, the value of design of experiment is a constant, i.e. $\delta_i = \frac{1}{n}$. Variable $w_i = \frac{z_i}{\sum_i zi}$ represents the above mentioned weight. Since this value is constant in formulas (2)–(5), it can be neglected, without any loss of generality, in subsequent calculations. Last, but not least, vector $\mathbf{f}_i$ is the variable which represents $i - th$ line of design matrix $\mathbf{F}$.

Expression $\frac{\sigma^2}{N}$ in (19) depends on the particular experiment, but it is the same for all models in (18). Thus, this expression can be put equal to 1, for the sake of comparison of formulas (2)–(5).

The comparison of regression models will be made by means of *a standard deviation* of individual unknown parameter estimates, i.e.

$$\mu(\widehat{\mathbf{\Theta}}) := \sqrt{\mathrm{diag}[\mathrm{Var}(\widehat{\mathbf{\Theta}})]}. \tag{21}$$

ad I) (one free parameter b)

Logarithmizing the equation (5), we arrive in view of (18) to the equality:

$$\begin{pmatrix} 0 \\ v_2 - \ln y_1 \\ \vdots \\ v_n - \ln y_1 \end{pmatrix} = \underbrace{\begin{pmatrix} -u_1 \\ \vdots \\ -u_n \end{pmatrix}}_{\mathbf{F}} \begin{pmatrix} b \end{pmatrix} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon}$ is an error vector with the covariance matrix equal to $\sigma^2\mathbf{I}$. The variables y_i, u_i and v_i have the same meaning as above.

The information matrix takes, according to (20) (when taking the weights $w_i = 1$), the form

$$\mathbf{M}(\delta) = \sum_{i=1}^{n} \frac{1}{n} (-u_i)(-u_i) = \frac{1}{n} \sum_{i=1}^{n} u_i^2,$$

from which (cf. (19), when taking $\frac{\sigma^2}{N} = 1$)

$$\mathrm{Var}_I(\widehat{\boldsymbol{\Theta}}(\mathbf{Y}_\delta)) \doteq 1.461694 \cdot 10^{-6},$$

and subsequently (see (6))

$$\mu(b_I) \doteq \sqrt{1.461694 \cdot 10^{-6}} \doteq 0.001209. \tag{22}$$

ad II) (two free parameters A, b)

Logarithmizing the equation (3), we arrive in view of (18) to the equality:

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & , -u_1 \\ \vdots & \\ 1 & , -u_n \end{pmatrix}}_{\mathbf{F}} \begin{pmatrix} a \\ b \end{pmatrix} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon}$ has the same meaning as in the foregoing case. The variables y_i, u_i and v_i have the same meaning as above and $a = \ln A$.

The information matrix takes, according to (20) (when taking the weights $w_i = 1$), the form

$$\mathbf{M}(\delta) = \sum_{i=1}^{n} \frac{1}{n} \begin{pmatrix} 1 \\ -u_i \end{pmatrix} \begin{pmatrix} 1, & -u_i \end{pmatrix}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1, & -u_i \\ -u_i, & u_i^2 \end{pmatrix} = \frac{1}{n} \begin{pmatrix} n, & -\sum_{i=1}^{n} u_i \\ -\sum_{i=1}^{n} u_i, & \sum_{i=1}^{n} u_i^2 \end{pmatrix},$$

from which (cf. (19), when taking $\frac{\sigma^2}{N} = 1$)

$$\mathrm{Var}_{II}(\widehat{\boldsymbol{\Theta}}(\mathbf{Y}_\delta)) \doteq \begin{pmatrix} 0.006875 & 0.000825 \\ 0.000825 & 0.000100 \end{pmatrix},$$

and subsequently (see (8), (9))

$$\mu(\ln A_{II}) \doteq \sqrt{0.006875} \doteq 0.082921, \tag{23}$$

i.e. (cf. [14, p. 215])

$$\mu(A_{II}) = \left.\frac{\partial e^a}{\partial a}\right|_{a=\ln A_{II}} \mu(\ln A_{II}) = A_{II}\mu(\ln A_{II}) \doteq 2.0815 \cdot 0.08291 \doteq 0.170919, \tag{24}$$

$$\mu(b_{II}) \doteq \sqrt{0.000100} \doteq 0.010025. \tag{25}$$

ad III) (two free parameters b, c)

Logarithmizing the equation (4), we arrive in view of (18) the equality:

$$\begin{pmatrix} 0 \\ v_2 - \ln y_1 \\ \vdots \\ v_n - \ln y_1 \end{pmatrix} = \underbrace{\begin{pmatrix} -u_1 & , x_1 - 1 \\ \vdots & \\ -u_n & , x_n - 1 \end{pmatrix}}_{\mathbf{F}} \begin{pmatrix} b \\ c \end{pmatrix} + \boldsymbol{\varepsilon}.$$

The vector $\boldsymbol{\varepsilon}$ and the variables y_i, u_i and v_i have the same meaning as above.

The information matrix takes, according to (20), (when taking the weights $w_i = 1$), the form

$$\mathbf{M}(\delta) = \sum_{i=1}^{n} \frac{1}{n} \begin{pmatrix} -u_i \\ x_i - 1 \end{pmatrix} \left(-u_i,\ x_i - 1 \right) = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} u_i^2, & -u_i(x_i - 1) \\ -u_i(x_i - 1), & (x_i - 1)^2 \end{pmatrix}$$

$$= \frac{1}{n} \begin{pmatrix} \sum_{i=1}^{n} u_i^2, & -\sum_{i=1}^{n} u_i(x_i - 1) \\ -\sum_{i=1}^{n} u_i(x_i - 1), & \sum_{i=1}^{n} (x_i - 1)^2 \end{pmatrix},$$

from which (cf. (19), when taking $\frac{\sigma^2}{N} = 1$)

$$\mathrm{Var}_{III}(\widehat{\boldsymbol{\Theta}}(\mathbf{Y}_\delta)) \doteq 1.0 \cdot 10^{-5} \begin{pmatrix} 0.868651 & 0.001135 \\ 0.001135 & 0.000001 \end{pmatrix},$$

and subsequently (see (11), (12))

$$\mu(b_{III}) \doteq \sqrt{0.868651} \doteq 0.002947, \tag{26}$$

$$\mu(c_{III}) \doteq \sqrt{0.000001} \doteq 0.000004. \tag{27}$$

ad IV) (three free parameters A, b, c)

Logarithmizing the equation (2), we arrive in view of (18) the equality:

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & , -u_1 & , x_1 \\ \vdots & & \\ 1 & , -u_n & , x_n \end{pmatrix}}_{\mathbf{F}} \begin{pmatrix} a \\ b \\ c \end{pmatrix} + \boldsymbol{\varepsilon}.$$

The vector $\boldsymbol{\varepsilon}$ and the variables y_i, u_i and v_i have the same meaning as above.

The information matrix takes according to (20), (when taking the weights $w_i = 1$), the form

$$\mathbf{M}(\delta) = \sum_{i=1}^{n} \frac{1}{n} \begin{pmatrix} 1 \\ -u_i \\ x_i \end{pmatrix} \left(1\ , -u_i\ , x_i \right) = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1, & -u_i, & x_i \\ -u_i, & u_i^2, & -u_i x_i \\ x_i, & -u_i x_i, & x_i^2 \end{pmatrix}$$

$$= \frac{1}{n} \begin{pmatrix} n, & -\sum_{i=1}^{n} u_i, & \sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} u_i^2, & \sum_{i=1}^{n} u_i^2, & -\sum_{i=1}^{n} u_i x_i \\ \sum_{i=1}^{n} x_i, & -\sum_{i=1}^{n} u_i x_i, & \sum_{i=1}^{n} x_i^2 \end{pmatrix},$$

from which (cf. (19), when taking $\frac{\sigma^2}{N} = 1$)

$$\mathrm{Var}_{IV}(\widehat{\mathbf{\Theta}}(\mathbf{Y}_\delta)) \doteq \begin{pmatrix} 0.018689 & 0.002724 & 0.75 \cdot 10^{-8} \\ 0.002724 & 0.000405 & 0.12 \cdot 10^{-8} \\ 0.75 \cdot 10^{-8} & 0.12 \cdot 10^{-8} & 0.48 \cdot 10^{-12} \end{pmatrix},$$

and subsequently (see (14)–(16))

$$\mu(\ln A_{IV}) \doteq \sqrt{0.018689} \doteq 0.136710, \tag{28}$$

i.e. (cf. [14, p. 215])

$$\mu(A_{IV}) = \left.\frac{\partial e^a}{\partial a}\right|_{a=\ln A_{IV}} \mu(\ln A_{IV}) = A_{IV}\mu(\ln A_{IV}) \doteq 2.2939 \cdot 0.13671 \doteq 0.313600, \tag{29}$$

$$\mu(b_{IV}) \doteq \sqrt{0.000405} \doteq 0.020148, \tag{30}$$

$$\mu(c_{IV}) \doteq \sqrt{0.48 \cdot 10^{-12}} \doteq 0.000007. \tag{31}$$

Summing up the above calculations, we can give the following theorem.

Theorem 2 *The standard deviation μ of parameters a, b, c in formulas (2)–(5) of MAL takes (by means of (19)–(21)) the respective forms (22) or (23), (25) or (26), (27) or (28), (30), (31). The deviations are universal in the sense that they are independent of concrete data $x_1, x_2, \dots; y_1, y_2, \dots; z_1, z_2, \dots$, and subsequently of concrete values of parameters b_I or A_{II}, b_{II} or b_{III}, c_{III} or A_{IV}, b_{IV}, c_{IV}, respectively. The deviations $\mu(A_{II})$ and $\mu(A_{IV})$ are no longer universal, but they satisfy the respective estimates (24) and (29).*

Remark 2 There is a question about the correct usage of regression models I) and III) in the above mentioned form. It concerns the fixed value y_1 which can be further regarded as a constant by means of which the remaining unknown parameters are estimated. Nevertheless, since the value y_1 represents the realization of a random variable, it contains the measurement error which is subsequently transferred to the regression model. Unfortunately, the regression model does not take this "double error" into account. By this reason, we must consider all the resulting values obtained by means of formulas (4) and (5) only as those conditioned by the value of y_1.

4 Tolerance region for neglecting parameter c

In this section, the influence of neglecting parameter c (i.e. $c = 0$) will be discussed in a more detail.

ad I and III) (two free parameters b, c vs. one free parameter b)

We consider the same regression model as in the previous chapter but, this time, we will write the plan matrix $\mathbf{F}$ in a different way, namely in the form of a block matrix. The aim is to separate the unknown parameter c from the remaining unknown parameters. Thus, we obtain the model given by

$$\begin{pmatrix} v_2 - \ln y_1 \\ \vdots \\ v_n - \ln y_1 \end{pmatrix} = (\mathbf{X}, \mathbf{S}) \begin{pmatrix} b \\ c \end{pmatrix} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon}$ is an error vector with the covariance matrix equal this time to $\sigma^2 \mathbf{V}^{-1}$, i.e.

$$\mathbf{X} = \begin{pmatrix} -u_2 \\ \vdots \\ -u_n \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} x_2 - 1 \\ \vdots \\ x_n - 1 \end{pmatrix}, \quad \mathbf{V}^{-1} = \begin{pmatrix} \frac{1}{w_2} + \frac{1}{w_1}, & \frac{1}{w_1}, & \dots & \frac{1}{w_1} \\ \frac{1}{w_1}, & \frac{1}{w_3} + \frac{1}{w_1}, & \dots & \frac{1}{w_1} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{w_1}, & \dots & \frac{1}{w_1}, & \frac{1}{w_n} + \frac{1}{w_1} \end{pmatrix}.$$

The variables y_i, u_i and v_i have the same meaning as those already discussed above. The matrix $\mathbf{V}$ is given by a concrete experiment (in our case, see Example 3). After performing a relatively long procedure (see [7, p. 95]), we arrive at the expression describing the *tolerance region*, i.e.,

$$\kappa = \left\{ c\colon |c| \le \frac{\sigma}{\sqrt{\mathbf{S}'\mathbf{V}\mathbf{X}\left\{\mathbf{X}'\mathbf{V}\mathbf{S}\left[\mathbf{S}'\left(\mathbf{N_X}\mathbf{V}^{-1}\mathbf{N_X}\right)^{+}\mathbf{S}\right]^{-1}\mathbf{S}'\mathbf{V}\mathbf{X}\right\}^{+}\mathbf{X}'\mathbf{V}\mathbf{S}}} \right\}, \tag{32}$$

where $\mathbf{N_X}$ is the projection matrix of the form

$$\mathbf{N_X} = \underbrace{\begin{pmatrix} 1, & 0, & \dots, & 0 \\ 0, & 1, & \dots, & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0, & \dots, & 0, & 1 \end{pmatrix}}_{\mathbf{I}} - \begin{pmatrix} -u_2 \\ \vdots \\ -u_n \end{pmatrix} \left(\sum_{i=2}^{n} u_i^2 \right)^{-1} \left(-u_2, \dots, -u_n \right).$$

The symbol "+" indicates the Moore–Penrose inverse (cf. e.g. [14, p. 40]). The tolerance region for a parameter c estimate depends on the value of σ. If we do not know the value of σ^2 from the previous measurements, then we must estimate it by the formula

$$\widehat{\sigma^2_{III}} = \frac{\left[y_* - (\mathbf{X}, \mathbf{S}) \begin{pmatrix} b_{III} \\ c_{III} \end{pmatrix} \right]' \mathbf{V} \left[y_* - (\mathbf{X}, \mathbf{S}) \begin{pmatrix} b_{III} \\ c_{III} \end{pmatrix} \right]}{n-k}, \tag{33}$$

where $y_* = (y_2, \dots, y_n)'$, k is the total number of unknown parameters in the model (4), i.e. $k = 2$, and n is the total number of measurement points in the tables in Example 1 without the value y_1, i.e. $n = 19$.

Example 5 Consider the same tables as in Example 1 and let us note that the analogy of Remark 1 holds here as well. In view of (33), we get that

$$\widehat{\sigma_{III}} = \sqrt{\widehat{\sigma^2_{III}}} \doteq 0.102319$$

and, according to (32), we arrive at $\kappa_{III} \doteq [-0.103574, 0.103574]$.

Since $c_{III} \doteq 0.0918$ in Example 3 satisfies $-0.103574 < 0.0918 < 0.103574$, i.e. $c_{III} \in \kappa_{III}$, the parameter c_{III} can be neglected in formula (4). In other words, formula (4) can be compensated by formula (5), in this case.

ad II and IV) (three free parameters A, b, c vs. two free parameters A, b)

We again rewrite the original regression model described above into the block form, i.e.

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = (\mathbf{X}, \mathbf{S}) \begin{pmatrix} a \\ b \\ c \end{pmatrix} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon}$ is again an error vector with the covariance matrix equal to $\sigma^2\mathbf{V}^{-1}$, i.e.

$$\mathbf{X} = \begin{pmatrix} 1, & -u_1 \\ \vdots & \vdots \\ 1, & -u_n \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{V}^{-1} = \begin{pmatrix} \frac{1}{w_1} & 0 & \dots & 0 \\ 0 & \frac{1}{w_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \frac{1}{w_n} \end{pmatrix}.$$

The variables y_i, u_i and v_i have the same meaning as above and the matrix $\mathbf{V}$ is given by a concrete experiment (in our case, see Example 4). The projection matrix $\mathbf{N_X}$ in the equation (32) takes now the form

$$\mathbf{N_X} = \mathbf{I} - \begin{pmatrix} 1, & -u_1 \\ \vdots & \vdots \\ 1, & -u_n \end{pmatrix} \begin{pmatrix} n, & -\sum_{i=1}^n u_i \\ -\sum_{i=1}^n u_i, & \sum_{i=1}^n u_i^2 \end{pmatrix}^{-1} \begin{pmatrix} 1, & \dots, & 1 \\ -u_1, & \dots, & -u_n \end{pmatrix}.$$

In Example 4, the value of σ can be estimated as follows:

$$\widehat{\sigma^2_{IV}} = \frac{\Delta_{IV}}{n-k} \;\Rightarrow\; \widehat{\sigma_{IV}} = \sqrt{\widehat{\sigma^2_{IV}}}, \tag{34}$$

where the value of Δ_{IV} comes from (17) and $n = 20$, in this case.

Example 6 Consider the same tables as in Example 1 and let us note that the analogy of Remark 1 holds here as well. In view of (34), we get that

$$\widehat{\sigma_{IV}} \doteq 0.079890$$

and, according to (32), we arrive at $\kappa_{IV} \doteq [-0.047844, 0.047844]$.

Since $c_{IV} \doteq 0.0442$ in Example 4 satisfies $-0.047844 < 0.0442 < 0.047844$, i.e. $c_{IV} \in \kappa_{IV}$, the parameter c_{IV} can be neglected in formula (2). In other words, formula (2) can be compensated by formula (3), in this case.

5 Concluding remarks

In Sections 2 and 3, the comprehensive analysis was done which provides a comparison of four formulas (2)–(5) from the point view of the best approximation of data as well as of the accuracy of parameters A, b, c estimates. These two approaches are based on the least square method, but provide somewhat different goals. Thus, we provide an information about a suitability of the usage of models which are evaluated from two perspectives.

As already pointed out, from the point of view of the first perspective, it might be expected that the best results can be obtained by means of formula (2), while the worst by means of formula (5). In order to compare formulas (3) and (4), containing two free parameters, one can immediately say only the following. Checking the inequalities (10) and (13), for $b_{II} \doteq b_{III}$, if $A_{II} < y_1$ and $c_{III} \geq 0$, then $\Delta_{II} < \Delta_{III}$ and, reversely, if $A_{II} > y_1$ and $c_{III} \leq 0$, then $\Delta_{II} > \Delta_{III}$. Otherwise, more sophisticated comparison techniques must be applied.

In view of the analysis in Sections 3 and 4, the situation is however much more delicate. In fact, since the accuracy of calculated parameters in formulas (2)–(5) is given by the inequalities $\mu(b_I) < \mu(b_{III}) < \mu(b_{II}) < \mu(b_{IV})$, the related errors must be balanced in an optimal way (whence the title of the paper).

For instance, in our linguistic experiment, since $\Delta_{III} < \Delta_{II}$ as well as $\mu(b_{III}) < \mu(b_{II})$, the application of formula (4) is evidently better than of (3). Nevertheless, because of $\Delta_{IV} < \Delta_{III} < \Delta_I$ and, at the same time, of $\mu(b_I) < \mu(b_{III}) < \mu(b_{IV})$, it is so far difficult to make any conclusion concerning the privileged formulas. This will be done later, on the basis of conclusions in Section 4.

Furthermore, although the shape of graph to functions associated with formulas (2)–(5) does not seem to differ much when taking into account the weights w_i (see Figures 1–4) or omitting them (see Figures 5–8), the respective least square values differ more than by one order. More precisely, the least square values including the weights are more than eleven times better than those with $w_i = 1$, $i = 1, 2, \ldots$

Replacing, for the same goal, the formulas (2)–(5) by polynomials of second (i.e. with three coefficients) or third (i.e. with four coefficients) degree (for the associated graphs, see Figures 9 and 10), the related least square values are $\Delta \doteq 0.1126$ and $\Delta \doteq 0.1123$. One can readily check that they do not only differ much each from other, but also from $\Delta_{III} \doteq 0.1234$ and $\Delta_{IV} \doteq 0.1085$. On the other hand, no graph in Figures 3, 4, 7, 8, 9 and 10 reflects, for higher values of $x_i = i$, the verbal form of MAL, because the associated functions are not decreasing.

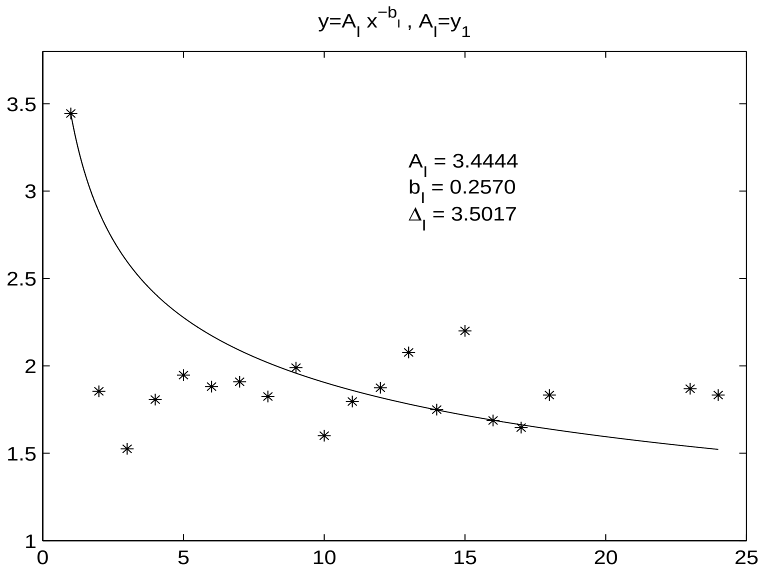

Figure 5: Calculated data y_i (see the tables above) versus the values $y(x_i) = 3.4444i^{-0.2570}$, with $w_i = 1$, $i = 1, 2, \ldots, 24$.

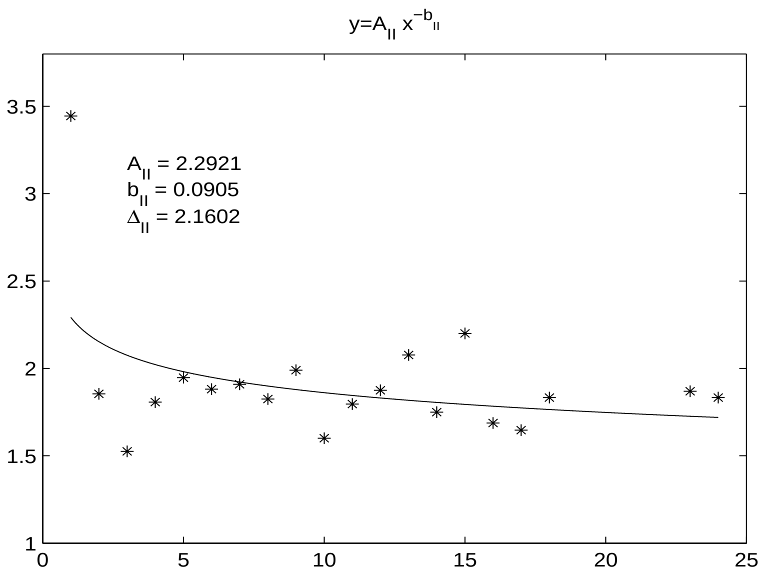

Figure 6: Calculated data y_i (see the tables above) versus the values $y(x_i) = 2.2921i^{-0.0905}$, with $w_i = 1$, $i = 1, 2, \ldots, 24$.

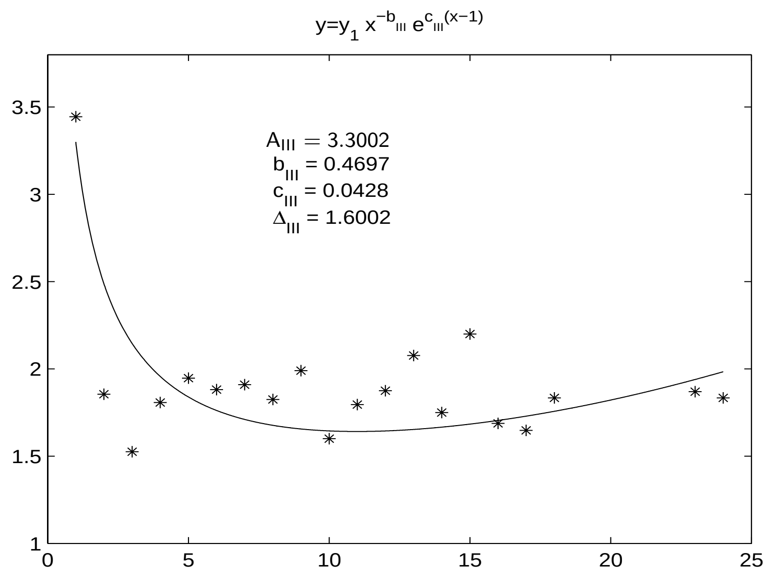

Figure 7: Calculated data y_i (see the tables above) versus the values $y(x_i) = 3.3002i^{-0.4697}e^{0.0428(i-1)}$, with $w_i = 1$, $i = 1, 2, \ldots, 24$. Since the related function is not decreasing, it becomes linguistically uninterpretable.

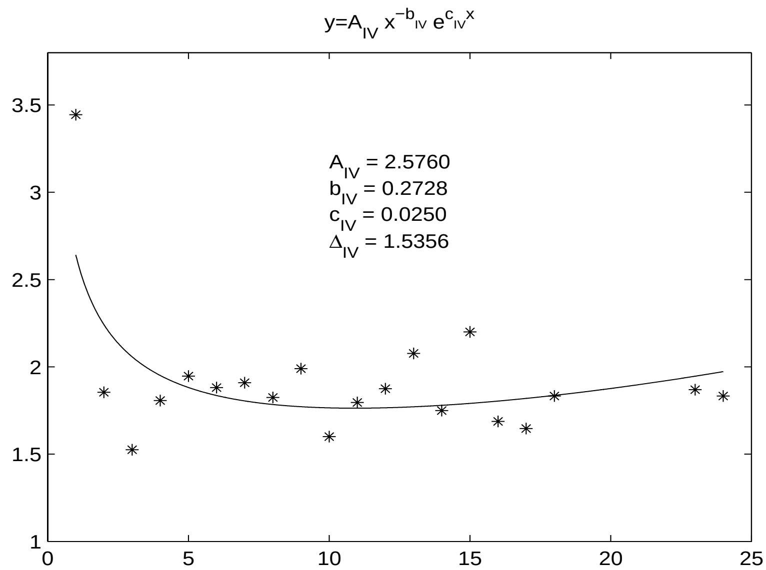

Figure 8: Calculated data y_i (see the tables above) versus the values $y(x_i) = 2.5760i^{-0.2728}e^{0.0250i}$, with $w_i = 1$, $i = 1, 2, \ldots, 24$. Since the related function is not decreasing, it becomes linguistically uninterpretable.

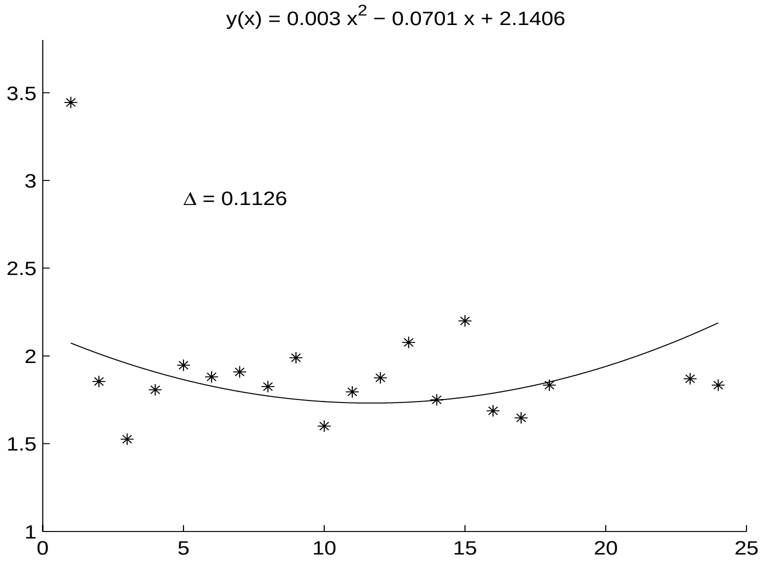

Figure 9: Calculated data y_i (see the tables above) versus the values $y(x_i) = 0.0030x_i^2 - 0.0701x_i + 2.1406$, $i = 1, 2, \ldots, 24$.

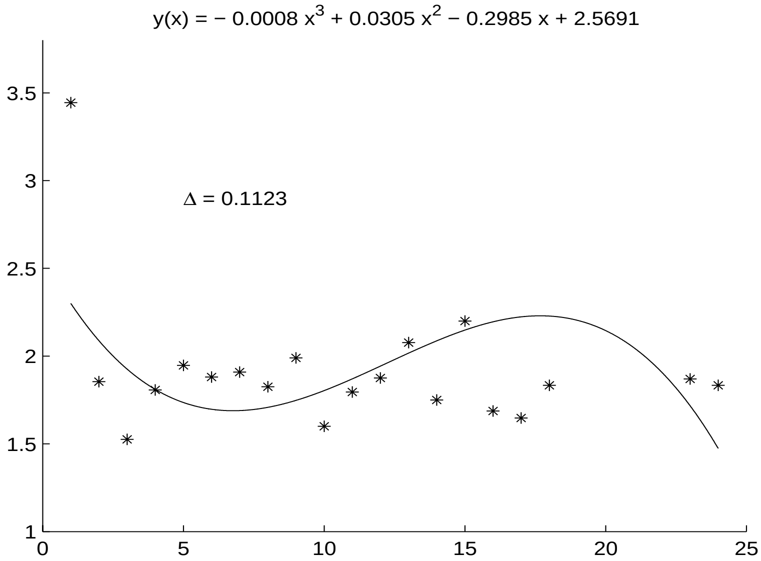

Figure 10: Calculated data y_i (see the tables above) versus the values $y(x_i) = -0.0008x_i^3 + 0.0305x_i^2 - 0.2985x_i + 2.5691$, $i = 1, 2, \ldots, 24$.

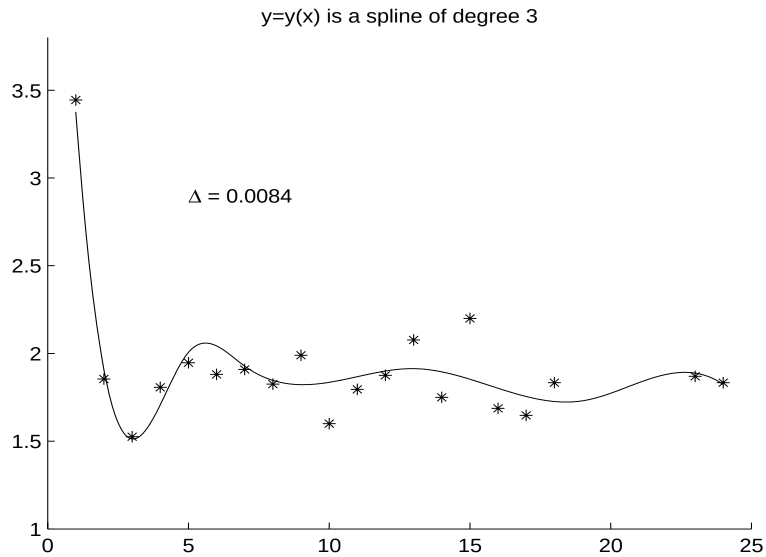

Figure 11: Calculated data y_i (see the tables above) versus the values $y(x_i) = -0.0808\,(x_i-1)^3 + 0.7790\,(x_i-1)^2 - 2.1664\,(x_i-1) + 3.3759$, for $x_i \in \langle 1,5\rangle$, $y(x_i) = 0.0380\,(x_i-5)^3 - 0.1901\,(x_i-5)^2 + 0.1893\,(x_i-5) + 2.0059$, for $x_i \in \langle 5,7\rangle$, $y(x_i) = -0.0031\,(x_i-7)^3 + 0.0377\,(x_i-7)^2 - 0.1156\,(x_i-7) + 1.9278$, for $x_i \in \langle 7,13\rangle$, $y(x_i) = 0.0023\,(x_i-13)^3 - 0.0187\,(x_i-13)^2 - 0.0020\,(x_i-13) + 1.9136$, for $x_i \in \langle 13,19\rangle$, $y(x_i) = -0.0049\,(x_i-19)^3 + 0.0232\,(x_i-19)^2 + 0.0248\,(x_i-19) + 1.7302$, for $x_i \in \langle 19,24\rangle$, $i = 1,2,\ldots,24$.

Applying, again for the same goal, a suitable spline of the third degree (see e.g. the one in Figure 11 with 20 coefficients), the related least square value can be very small (in our case, $\Delta \doteq 0.0084$). Although, the associated function is not decreasing, one can say that "statistically", or "in the average", it has already a decreasing tendency, in accordance with the verbal form of MAL. On the other hand, this is due to the quite noneffective presence of 20 coefficients.

From the point of view of the second perspective, besides the mentioned inequalities

$$\mu(b_I) < \mu(b_{III}) < \mu(b_{II}) < \mu(b_{IV}), \tag{35}$$

we can also write that

$$(\mu(a_{II}) < \mu(a_{IV}) \Rightarrow)\, \mu(A_{II}) < \mu(A_{IV}), \quad \text{and} \quad \mu(c_{III}) < \mu(c_{IV}).$$

According to the analysis in Section 4, the complete formula (2) can be compensated by the truncated one (3), as well as (4) by (5). In this light and because of (35), formula (5) is in fact better than (4). On the other hand, despite the fact that (2) can be compensated by (3) and the inequalities $\mu(b_{II}) < \mu(b_{IV})$, in view of parameter values $b_{II} \doteq 0.0753$ and $b_{IV} \doteq 0.3094$, formula (3) is the worst of all. After all, since (4) was already shown to be more suitable than (3), the simplest formula (5) is optimal for the computation of parameter b. Thus, $b_I \doteq 0.3574$ is the best estimate (among $b_I, b_{II}, b_{III}, b_{IV}$) of parameter b, in our linguistic experiment. We believe that this particular case can be extended to a more general situation which will be treated by ourselves elsewhere.

References

[1] Altmann, G.: *Prolegomena to Menzerath's law.* Glottometrika 2, **2**, (1980), 1–10.

[2] Altmann, G., Schwibbe, M. H. (eds.): *Das Menzerathsche Gesetz in informations – verarbeitenden Systemen. Olms*, Hildesheim, 1989.

[3] Andres, J.: *On de Saussure's principle of linearity and visualization of language structures.* Glottotheory **2**, 2, (2009), 1–14.

[4] Andres, J.: *On a conjecture about the fractal structure of language.* Journal of Quantitative Linguistics **17**, 2, (2010), 101–122.

[5] Andres, J., Benešová, M.: *Fractal analysis of Poe's Raven.* Glottometrics **21** (2011), 73–98.

[6] Cramer, I.: *The parameters of the Altmann–Menzerath law.* Journal of Quantitative Linguistics **12**, 1,(2005), 41–52.

[7] Fišerová, E., Kubáček, L., Kunderová, P.: Linear Statistical Models. Regularity and Singularities. *Academia*, Prague, 2007.

[8] Hřebíček, L.: *The constants of Menzerath–Altmann's law.* Glottometrika. **12** (1990), 61–71.

[9] Hřebíček, L.: *Text Levels. Language Constructs, Constituents and the Menzerath–Altmann Law. Wissenschaftlicher Verlag Trier*, Trier, 1995.

[10] Hřebíček, L.: *Text in Semantics. The Principle of Compositeness. Oriental Institute of the Academy of Sciences of the Czech Republic*, Prague, 2007.

[11] Köhler, R.: *Das Menzerathsche Gesetz auf Satzebene.* Glottometrika **4** (1982), 103–113.

[12] Köhler, R.: *Zur Interpretation des Menzerathschen Gesetzes.* Glottometrika **6** (1984), 177–183.

[13] Köhler, R.: *Das Menzerathsche Gesetz als Resultat des Sprachverarbeitungsmechanismus.* In: Altmann G. und Schwibbe, M. H. (eds.): *Das Menzerathsche Gesetz in informations – verarbeitenden Systemen. Olms*, Hildesheim, 1989, 108–116.

[14] Kubáček, L., Kubáčková, L.: Statistics and Metrology. *Palacký University*, Olomouc, 2000, (in Czech).

[15] Kulacka, A.: *The coefficients in the formula for the Menzerath–Altmann law.* Journal of Quantitative Linguistics **17**, 4, (2010), 257–268.

[16] Kulacka, A., Mačutek, J.: *A discrete formula for the Menzerath–Altmann law.* Journal of Quantitative Linguistics **14**, 1 (2007), 23–32.

[17] Pázman, A: Foundations of Optimum Experimental Design. *D. Reidel*, Dordrecht, 1986.

[18] Prün, C.: *Validity of Menzerath-Altmann's law: graphic representation of language, information processing systems and synergetic linguistics.* Journal of Quantitative Linguistics **1**, 2 (1994), 148–155.

[19] Tučková, M: *Optimal Design of Regression Experiments.* Diploma Thesis, *Palacký University*, Olomouc, 2010 (in Czech).

[20] Wimmer, G., Altmann, G.: *Unified derivation of some linguistic laws.*. In: Köhler R., Altmann G., Piotrowski R. G.: Quantitative Linqustics. An International Handbook. *De Gruyter*, Berlin, 2005, 791–807.

[21] Wimmer, G., Altmann, G., Hřebíček, L., Ondrejovič, S., Wimmerová, S.: *Introduction to the Analysis of Texts. Veda*, Bratislava, 2003, (in Slovak).