Practical View on Calibration

Assoc. prof. Milan Javůrek[1] and Prof. Ivan Taufer

Department of Process Control,
Faculty of Electrical Engineering and Informatics,
University of Pardubice,
Nám. Čs. Legií 565, 532 10 Pardubice,
Czech Republic
E-mail: Milan.Javurek@upce.cz; Ivan.Taufer@upce.cz
Tel.:+420 466 037 103 or +420 603 443 819

Abstract
Calibration is one of the most frequently used laboratory procedures. Thanks to immense development of measuring techniques at present, the measurement procedures have become easier both quantitatively and qualitatively, i.e. the precision and correctness of measurements and their speed have increased; hence the amount of experimental data has multiplied. Along with experimental techniques, the techniques of evaluation have also improved: available computer programs cover absolute majority of mathematical evaluation methods. In the profusion of outputs, the application conditions of the methods used are somewhat eluding, and the results obtained often have limited applicability. Therefore, it is necessary to become aware of the fact that for practical use the methods have to be clear, straightforward and exact, i.e. it has to be known what can be neglected as marginal and what cannot.

Key words: calibration, calibration curve, regression analysis, model searching.

1. Introduction

The term calibration itself has a somewhat wider meaning. Very erudite, even though rather theoretical view of calibration is offered in (Vdoleček et al., 2002). In metrological sense, calibration represents the adjustment of measuring instruments on etalon (and/or etalons, as the case may be) with the aim of ensuring qualitative correctness of measurement (e.g., adjusting of pH-meter by means of standard buffers). According to the procedure adopted, we distinguish comparison methods with one standard, when the experimental conditions during "standardization" and measurement should be identical (e.g., concentration, ionic strength etc.). This condition, however, tend to not be fulfilled, and the problem in the case of chemical analyses can be dealt with by the method of standard addition, when after measuring an unknown sample we add a known amount of the substance and again measure the response. That, however, is not viable in a number of experiments (branches). The most common task used is the so-called ***absolute***

[1] Corresponding author.

calibration. In this calibration we measure, for a selected range of independent variable x, the required experimental dependence $y = f(x)$. Then for the measured value of dependent variable y of an unknown sample we are able to find the looked for (unknown) value of the quantity x (e.g., concentration) of this sample. During this procedure, we suppose that the values of independent variable are not loaded with an error, whereas random errors can exist among the measured values. Other types of errors have to be eliminated (e.g., by means of the above-mentioned metrological calibration).

2. Determination of Model and Test of Remoteness of Points

The introductory step is the measurement of calibration dependence. The experiment is realized in such a way as to have the values of independent variable which as far as possible symmetrically cover the region specified by the required extent of measurement. In this way we obtain a set (matrix) of data with the magnitude of $(m + 1)\cdot n$, where m is the number of independent variables, n is the number of experiments (patterns), which are geometrically presented as points in an $(m + 1)$-dimensional space. In some cases (branches) there may occur a problem connected with the fact that with regard to measurement possibilities (available materials, reproducibility etc.), the particular values of independent variable tend to be specified in branch regulations, standards etc. Unfortunately, these directions are not always elaborated in good quality and by well-informed authors. As an example, we can adduce the statistical standards ISO 5725.

The second step is fitting of the obtained experimental results (experimental points) with the chosen mathematical (graphical) dependence. There are two possible ways to deal with this:

- ***regression analysis*** of experimental dependence, most often by application of the chi-square test and its modifications. Here we have the possibility to replace the experimental dependence by a function that is linear or non-linear in parameters, i.e. to use linear or non-linear regression. The advantage is a smooth or smoothed functional dependence which compensates errors of the measured quantity;
- ***interpolation*** by means of interpolation polynomials, spline functions of different orders etc. However, in this case the chosen function "goes through" the experimental points, and possible marked errors of measurement can thus more or less distort the course of functional dependence.

At present, the given algorithms possess developed programs and are included into a number of mathematical and statistical computing systems. It depends on the user, who is expected to select the most suitable processing algorithm for dealing with a particular task: an algorithm which can ensure not only speed but also faultlessness and adequateness of the solution.

From the logic of this matter follows that in the case of absolute calibration we can request that the calibration function should have a smooth, monotonic and not very complicated course. In the given case, the method of linear regression appears to be most suitable. Its further advantages include the following:

- simple and univocal calculation course, its availability in all program systems designed for evaluation of experimental data;
- by using various functions with linear parameters it is possible to approximate most necessary dependences;
- the linear regression has a very well elaborated system for calculation of parameter errors, properties of residua, identification of multicollinearity, correlation of variables, partial regression graphs etc.;
- there are available universal tools for reliable evaluation and comparison of fitting with various functions (Akaike's Information Criterion - AIC, mean quadratic error of prediction - MEP);
- effective identification of outlier and extreme points by application of visual graphical tools (graphical methods). Programs have been compiled for most of them, and they have been included into the corresponding computing systems. In particular, these involve the following:

- the graph of predicted residua
- Williams's graph
- Pregibon's graph
- McCulloh–Meeter's graph
- L – R graph;

Note. The basic principle of these methods lies in a defined (according to the method) transformation of coordinate values of experimental points and their placement on the corresponding coordinate surface. The surface is divided by means of dividing lines into segments of points, extreme points, and active points. Thereby the experimental points are divided according to these respective properties. It is recommended to eliminate outlier points from further processing, as they are points loaded with large error. Extreme points only signal a larger distance from the cluster (concentration) of active points; however, they are not eliminated from the processing. For more details of the methodology of this identification, see (Meloun and Militký, 2011). Practical examples are given below in chapter 6.

- calculation of the distances of individual points from the regression line (Cook's, Atkinson's, likelihood distance) and, in this way, evaluation of their effect upon the quality of fit (Meloun and Militký, 2011).

Univocally the most preferred model is the linear function, i.e. the calibration straight line in the following form

$$y = b_0 + b_1 x \qquad (1)$$

There exist dependences, such as the Lambert–Beer law, which are linear in the commonly used concentration range. If the requirements for the calibration range allow, then another possibility lies in the use of a linear part of the respective dependence. However, this is connected with a problem: which points can still be included as "linear" and which cannot. One of possible solutions to this problem is suggested in (Eckschlager et al., 1980):

"It is required that the product of deviations (residua) of two neighboring points from the fitted straight line should be ten times smaller that the sum of squares of the deviations of all the points (i.e. the fitness function) divided by their number."

Another possibility lies in the application of the already mentioned diagnostic graphs.

One of the mistakes that have been historically handed down is the application of various transformations that lead to a linearization of experimental dependence (e.g., transformation to logarithm). In this case, however, one of the basic requirements for application of regression method – the normal distribution of errors – is not fulfilled. The measurement has been distorted by the transformation. Apart from this precondition for application of the chi-square test there are several others:

- constant dispersion of measurement errors over the whole course of the dependence (homoskedasticity),
- the measurement errors have zero mean value and are not mutually correlated,
- the function (model) is linear in parameters, which can assume any arbitrary values; the only limitation exist in physical sense,
- the values of independent variable are not arbitrary, they can be adjusted.

Another fact that is very often ignored is the requirement for ensuring of sufficient number of the so-called degrees of freedom, i.e. the number of experimental data that can be used for one parameter determined. In the present authors' experience, this number should not sink below 3, the optimum value being 5.

3. Vagueness of Regression Function

With regard to the fact that the dependent variable, i.e. the measured quantity, is loaded with a random error, the individual experimental points should be symmetrically scattered around the regression dependence. This vagueness is expressed by the interval of prediction reliability, i.e. the confidence belt around the regression line, in which the experimental points are present with a chosen probability [2], and which is expressed as follows:

$$\boldsymbol{x}_0^T\ \boldsymbol{b} - t_{1-\alpha/2}\,(n-m)\,\hat{s}_{P,0} \le \boldsymbol{x}_0^T \boldsymbol{\beta} \le \boldsymbol{x}_0^T\ \boldsymbol{b} + t_{1-\alpha/2}\,(n-m)\,\hat{s}_{P,0} \tag{2}$$

where n – number of points

m – number of parameters

$\boldsymbol{x}_0^T$ – vector of values of independent variable of a given point

$\boldsymbol{b}$ – vector of estimates of parameters

$\boldsymbol{\beta}$ – vector of correct values of parameters

$t_{1-\alpha/2}$ – quantile of Student's distribution for a chosen significance level

$\hat{s}_{P,0}$ – relative standard deviation of prediction, whose square (e. g. dispersion) is expressed as:

$$D_{(\hat{y}_{P,0})} = \hat{s}_{P,0}^2 = \hat{\sigma}\ \boldsymbol{x}_0^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}_0 \tag{3}$$

where $\boldsymbol{X}$ – matrix of values of independent variables

$\hat{\sigma}$ – relative standard deviation of vector of independent variable

Thus Eq. (2) specifies the 100(1-α)% interval of prediction reliability $\hat{y}_{P,i}$ in the place x_0. The confidence belts are most narrow in the center of gravity of the independent variable(s) $x_{0j} = \bar{x}_j$.

If we need to determine the confidence belts for all possible values of the vector $\boldsymbol{x}$, then we use Scheffé's method. Then the value $\boldsymbol{x}^T \boldsymbol{\beta}$ lies in the interval

$$\boldsymbol{x}^T \boldsymbol{\beta} \pm \sqrt{m F_{1-\alpha}(m, n-m)}\ \hat{\sigma} \sqrt{\boldsymbol{x}^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}} \tag{4}$$

These belts are usually wider and are called Working–Hottelling belts in literature. However, the presented solution to the problem only applies to linear models.

The sense of calibration is finding of the value $\hat{x}^*$ corresponding to the relevant measured quantity y^*. As follows from the above statements, even for a linear model it is impossible to carry out a direct reading of independent variable – see the next chapter. For the most usual model of polynomial, this value is looked for as the root of the polynomial. An even more complicated situation is encountered in the case of non-linear regression, where the solution is looked for in the form

$$\hat{x}^* = f^{-1}(y^*) \tag{5}$$

For a more detailed description of the derivation, see (Meloun and Militký, 2011).

The principle of reading from calibration graph is represented in Fig. 1, (Meloun and Militký, 2011).

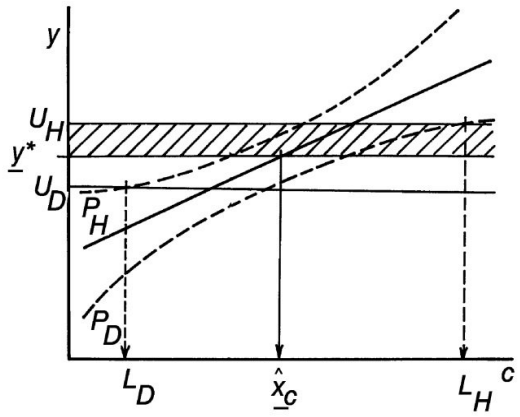


Fig. 1 – Determination of reliability interval $L_{D,H}$ of quantity x_C from calibration straight line. Hachuring denotes the half-width of reliability interval of signal. $P_{D,H}$ are the lower and upper parabolas of confidence belt, $U_{D,H}$ are the lower and upper confidence straight lines of measured quantity

The determination of reliability interval of independent variable represents a task of finding of points of intersection of the upper limit of reliability interval U_H with the lower parabola of the Working–Hottelling belt, and that of the lower limit U_D with the upper parabola.

4. Calibration Table

The calibration table in its core represents a record of the required or measured values of dependent variable and the corresponding values of independent variable calculated for a linear dependence from the inversive relation

$$\hat{x}^* = \bar{x} + \frac{y^* - \bar{y}}{b_1} \tag{6}$$

where $\bar{x}$, $\bar{y}$ are arithmetical means of the quantities,

y^* – the measured value of dependent variable,

b_1 – the angular coefficient of calibration straight line (the first parameter of the calibration straight line).

Due to the vagueness of the regression straight line, this estimate is deflected. Therefore, modified estimates are used:

1) *Naszodi's modified estimate*

$$\hat{x}_B^* = \bar{x} + \frac{(y^* - \bar{y})\, b_1}{b_1^2 + \dfrac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{7}$$

2) *Krjučkov's inversive estimate*

$$x_I^* = \bar{x} + (y^* - \bar{y}) \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{8}$$

3) *Schwartz's non-linear estimate*

$$x_N^* = \frac{\sum_{i=1}^{n} x_i \exp\left[\frac{-(y^* - b_0 - b_1 x_i)^2}{2\sigma^2}\right]}{\sum_{i=1}^{n} \exp\left[\frac{-(y^* - b_0 - b_1 x_i)^2}{2\sigma^2}\right]} \tag{9}$$

where b_0 is the second parameter of the calibration straight line.

The calibration table also includes the interval estimates described in the previous chapter. The table thus serves the presentation of calculation results in the course of calibration curve and its statistical parameters. An example of calibration table is given in chapter 6 (Table 2).

5. Calibration Accuracy

An important part of calibration calculations is the determination of the limit values of validity of calibration, which in the first place are connected with the existence of the so-called noise. In this case, the limits are defined within which the measured signal can be safely distinguished from the noise; they involve the following:

a) *critical level* y_C

It represents the minimum measured value which can still be distinguished from the noise; it is often called *blank measurement*

$$y_C = \bar{y} - b_1\bar{x} + \hat{\sigma}\, t_{1-\alpha}(n-2)\sqrt{1 + \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{10}$$

b) *detection limit* y_D

It corresponds to the value of dependent variable, where the critical level intersects with the lower parabola of the confidence belt of regression line. Actually, it is the critical level corrected by means of vagueness of regression line

$$y_D = y_C + \hat{\sigma}\, t_{1-\alpha}(n-2)\sqrt{1 + \frac{1}{n} + \frac{(x_D - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{11}$$

The principle is expressed in Fig. 2, (Meloun and Militký, 2011).

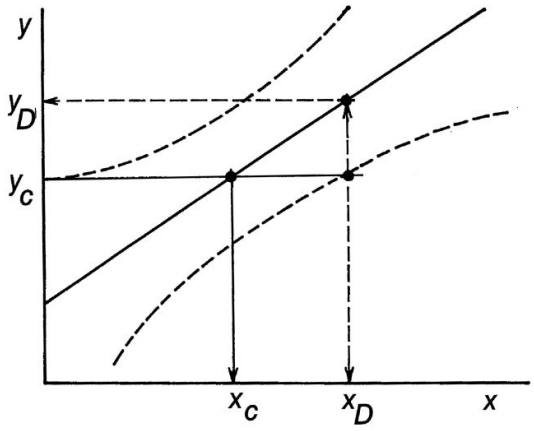


Fig. 2 – Representation of critical level y_C, detection limit y_D and the corresponding values x_C a x_D.

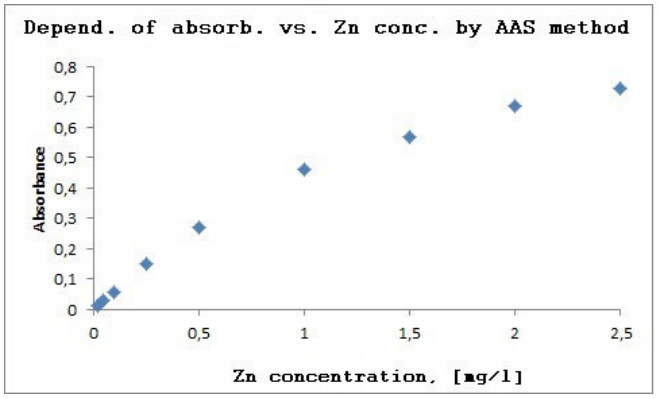
c) *determination limit* y_S

It is the smallest value of signal for which the relative standard deviation of prediction from the calibration model is equal to the preselected value *C*, usually 10%. For practical purposes, it is calculated from the formula

$$y_S = \frac{\hat{\sigma}}{C}\sqrt{1+\frac{1}{n}+\frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \qquad (12)$$

6. Example

For illustration of processing of calibration data, we have used assignment K6.03 from (Meloun and Militký, 2011). There were 9 experimental data available. Their graphical representation is given in Fig. 3.



6.1 Selection of model

Figure 3 shows that the course of experimental dependence is visually distinctly non-linear; hence it is useless to test the straight-line model. We will limit our selection to the second-order and third-order polynomials.

6.2 Test of remoteness of data

The tests of remoteness of data were performed by means of the above-mentioned graphical methods whose computer programs are included in the Adstat system (Meloun and Militký, 2011). The results of tests obtained by means of the individual methods are as follows:

Graphs of predicted residuals, Fig. 4 – for the polynomial of the 2nd order, the extreme points are No. 7 and 9; for the polynomial of the 3rd order, the remote point is No. 3, and the extreme points are No. 7, 8 and 9.

Pregibon graphs, Fig. 5 – all the points are active.

Williams's graphs, Fig. 6 – extreme point is only No. 9 for the 2nd order polynomial; for the 3rd order polynomial, outlier points are No. 7 and 8; No. 9 is extreme point.

McCulloh and Meeter graphs, Fig. 7 – for the 2nd order polynomial, No. 9 is extreme point; No. 6 and 7 can be classified as extreme points and outlier points; for the 3rd order polynomial, No. 7 and 8 are outlier points, No. 9 is extreme point.

L – R graphs, Fig. 8 – for the 2nd order polynomial, No. 7 is outlier point, No. 9 is extreme point; for the 3rd order polynomial, No. 7 and 8 are outlier points, No. 9 is extreme point.

From the above presentation follows that the 3rd order polynomial provides a somewhat closer fit, hence it declares more remote points, which is documented by at least three graphs. If these points were eliminated, the number of degrees of freedom for regression calculation would be lost. After the evaluation it can be concluded that the 2nd order polynomial appears to be more suitable (and/or sufficient, as the case may be).

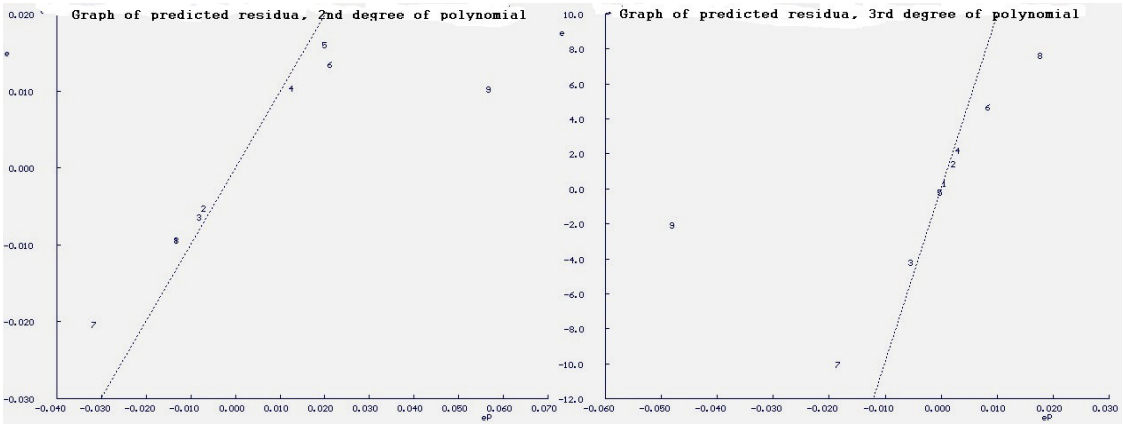


Fig. 4 – Graphs of predicted residuals for the 2nd and 3rd order polynomials

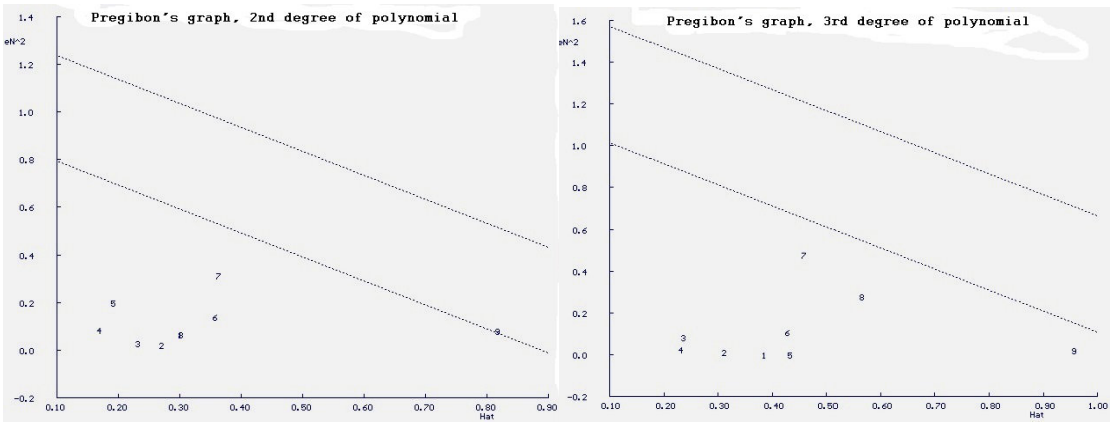


Fig. 5 – Pregibon graphs for the 2nd and 3rd order polynomials

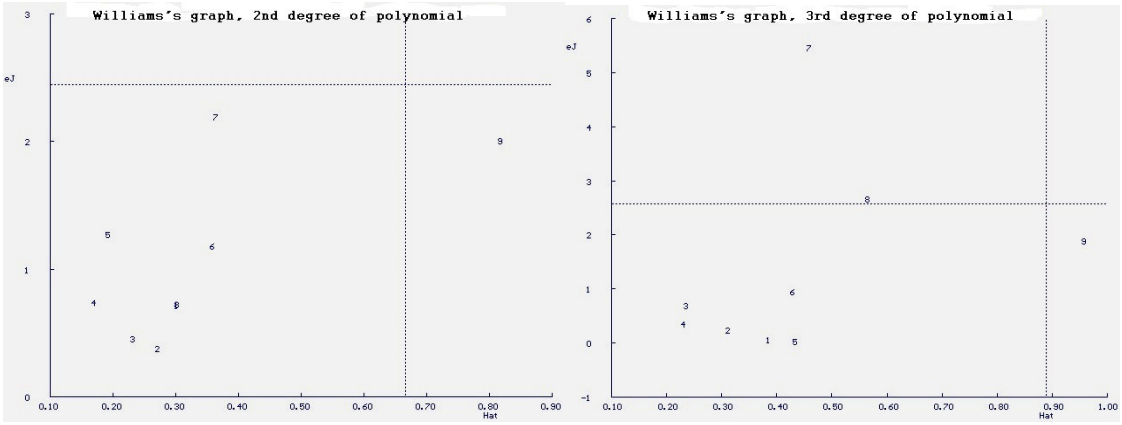


Fig. 6 – Williams graphs for the 2nd and 3rd order polynomials

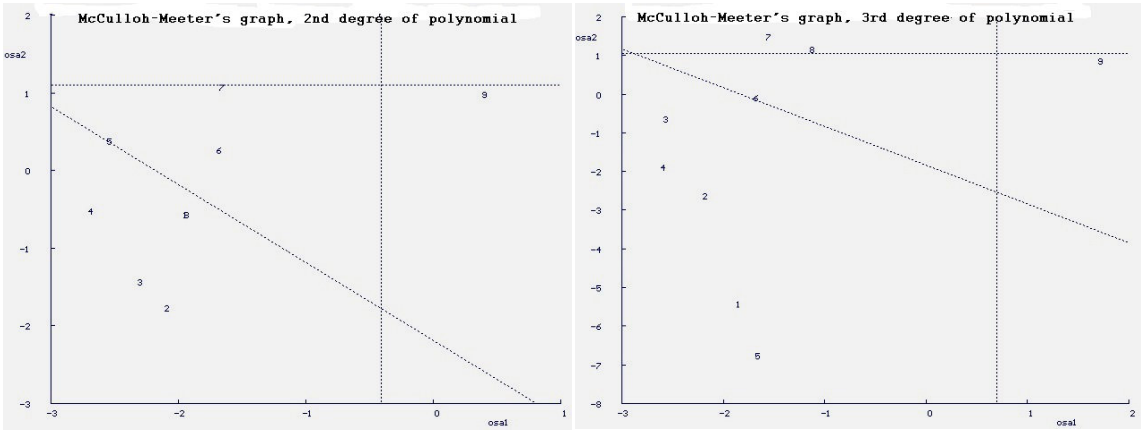


Fig. 7 – McCulloh and Meeter graphs for the 2nd and 3rd order polynomials

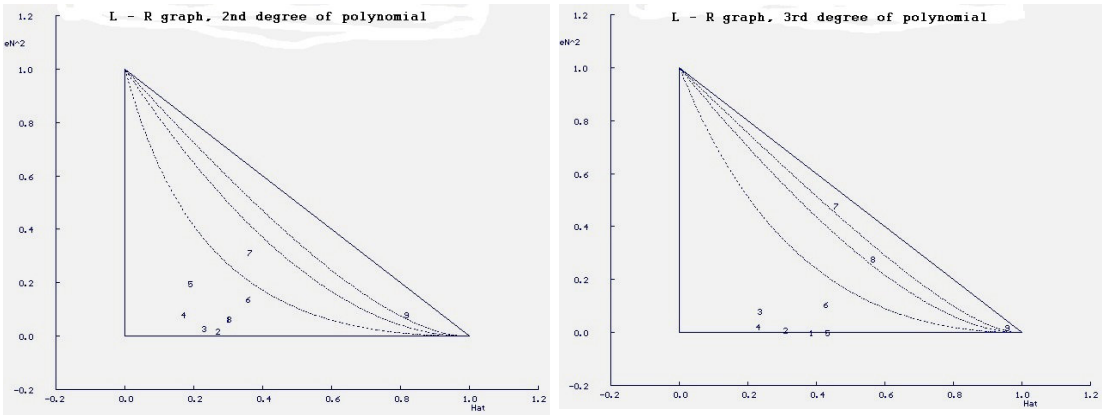


Fig. 8 – L–R graphs for the 2nd and 3rd order polynomials

6.3 Statistical characteristics

Table 1 presents the parameters and statistical characteristics calculated by regression for the 2nd and 3rd order polynomials. Here again the 2nd order polynomial appears to be sufficient, first of all due to the insignificant parameter No. 1 in the 3rd order polynomial and its large error; the other characteristics are broadly comparable. Hence an addition of further parameter brings no significant advantage.

Table 1 – Parameters and characteristics of the model looked for
(the values in round brackets are errors of parameters)

Characteristics	2nd order polynomial	3rd order polynomial
Parameter No. 1	$1.061 \cdot 10^{-2}$ (±0.837)	$-1.147 \cdot 10^{-3}$ (±4.337)
Parameter No. 2	$5.373 \cdot 10^{-1}$ (±0.216)	$6.447 \cdot 10^{-1}$ (±0.231)
Parameter No. 3	$-1.015 \cdot 10^{-1}$(±0.089)	$-2.202 \cdot 10^{-1}$ (±0.235)
Parameter No. 4	–	$3.184 \cdot 10^{-2}$ (±0.623)
t - test of significance of parameters	All are significant	Parameter No. 1 is insignificant
Mean quadratic error of prediction, MEP	$6.314 \cdot 10^{-4}$	$3.423 \cdot 10^{-4}$
Akaike's information criterion, AIC	$-7.352 \cdot 10^{1}$	$-8.797 \cdot 10^{1}$
Residual sum of squares, RSC	$1.309 \cdot 10^{-3}$	$2.105 \cdot 10^{-4}$
Heteroskedasticity of residua	yes	no
Normality of residua	yes	yes
Autocorrelation of residua	no	yes

6.3 Calibration table

Table 2 summarizes the results of the calibration itself with the model of the 2nd order polynomial. In assignment K 6.03, four measured values are assigned, whose value of independent variable is to be determined. The first of the values is smaller than the critical value; hence it cannot be taken into consideration during evaluation. If we compare the value of estimate with limits of reliability intervals, the non-linearity of the dependence will make itself felt in a non-symmetrical placement of the estimate inside the interval.

The course of calibration model inclusive of Working–Hottelling's belts is presented in Fig. 9.

Conclusion

As it was stated in Introduction, the technique of absolute calibration of linear (in parameters) dependences is used in laboratories very often. This is due to the fact that the method is simple in calculation, reliable and effective. In addition there already exist a number of computation systems (e.g., Adstat or its newer version QCExpert (Meloun and Militký, 2011)), which contain program modules necessary for the solution.

Table 2 – Results of calibration calculations for assignment K 6.03
(program QC Expert, (Meloun and Militký, 2011))

Critical level	y_C	x_C	
	$3.908 \cdot 10^{-2}$	$5.354 \cdot 10^{-2}$	
Detection limit	y_D	x_D	
	$6.324 \cdot 10^{-2}$	$9.985 \cdot 10^{-2}$	
Measured value y^*	Inversive estimate $\hat{x}^*$	Reliability interval $L_{D.H}$	
$5.00 \cdot 10^{-3}$	$-1.042 \cdot 10^{-2}$	$-6.963 \cdot 10^{-2}$	$3.936 \cdot 10^{-2}$
$1.55 \cdot 10^{-1}$	$2.839 \cdot 10^{-1}$	$2.410 \cdot 10^{-1}$	$3.266 \cdot 10^{-1}$
$3.55 \cdot 10^{-1}$	$7.461 \cdot 10^{-1}$	$6.817 \cdot 10^{-1}$	$8.192 \cdot 10^{-1}$
$5.55 \cdot 10^{-1}$	1.365	1.250	1.487

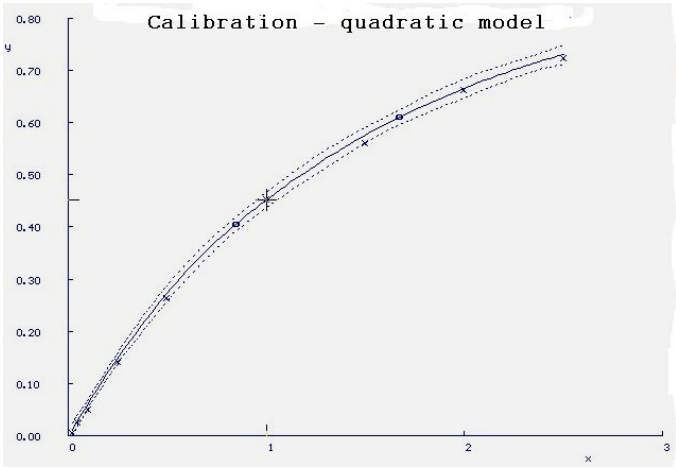


Fig. 9 – Resulting calibration model for assignment K 6.03

The application of this method presupposes first of all measurement of sufficient number of necessary data, with minimizing of system and gross errors of measurement if possible: this ensures a sufficient number of degrees of freedom. Therewith connected is also the selection of corresponding model(s). For each suggested model, the model parameters are calculated by means least square method, their statistical significance is tested, the calibration limits are calculated, and the outlier points are determined with the help of graphical tools of exploratory analysis. According to the results thus obtained, it is possible to decide which version of the suggested model is the best. Furthermore, it is possible to correct the data (by elimination of outlier points) and to reduce statistically insignificant parameters and then repeat the whole calculation process.

On the basis of the obtained calibration model, the calibration table for experimental values of dependent variable is compiled as necessary.

References

ECKSCHLAGER, K.; HORSÁK, I.; KODEJŠ, Z. 1980, *Evaluation of Analytical Results and Methods* (in Czech). Prague: SNTL, 1980. 224 pp.

MELOUN, M.; MILITKÝ, J. 2011, *Statistical Data Analysis. A practical guide.*, Woodhead Publishing Ltd. ISBN 978-0-85709-109-3. Including CD.

VDOLEČEK, F.; PALENČÁR, R.; HALAJ, M. 2002, Uncertainties in Measurement IV.: Uncertainties in Calibration (in Czech). *Automa, 8*, No. 4, 2002, pp. 41-47, ISSN 1210-9592.