

TREATMENT OF MODEL ERROR IN CALIBRATION  
BY ROBUST AND FUZZY PROCEDURES

KEY WORDS: Calibration, Model error, Robust and Fuzzy regressions

Róbert Rajkó

Department of Inorganic and Analytical Chemistry,  
A. József University, H-6701 Szeged, P.O. Box 440, Hungary

ABSTRACT

In every mathematical (e.g., statistical) procedure and theorem used in calibration, several conditions need to be fulfilled. What can analysts and chemometricians do, however, if the conditions are only nearly fulfilled? One can expect that small changes in the conditions yield only small changes in the results. This article shows how to treat two types of model error caused by assuming an incorrect error distribution or relationship (i.e., linear). The procedures applied are based on robust statistics and fuzzy theory, respectively.

INTRODUCTION

In calibration, it is important to determine the parameters of the calibration function by means of certain parameter estimators. But where can these estimators be derived from? It must first be defined what the calibration function means, i.e., what the shape of the function is. In practice, a linear one is often chosen for simplicity. We must then consider the error distribution. In general it may be assumed that the independent variable (concentration) is precise, i.e., it is free from any measurement

error (or at least its error is negligible as compared with that in the response). Accordingly we have to assume an error distribution only in the dependent variable (response). Finally, we can give a model for the calibration procedure with calibration function  $C(\cdot)$ :

$$\eta = C(x, \underline{p}) + \delta \quad (1)$$

where  $\delta$  is the measurement error in the response,  $\underline{p}$  is the parameter vector and  $\eta$  is a random variable related to the response.

In this paper we distinguish two types of model error:

- those caused by a difference from the assumed error distribution,
- those caused by a difference from the assumed class of the calibration function (i.e., from its linearity).

In the following sections we show how certain parameter estimators can be derived to treat the above model errors. As solution of the combination of two model errors a robust fuzzy method was introduced for artificial<sup>1</sup> and real<sup>2,3</sup> data.

#### MODEL ERROR CAUSED BY DIFFERENCE FROM THE ASSUMED ERROR DISTRIBUTION

We assume here that the calibration function is exact, but that the measurements from the analytical instrument can be obtained only contaminated with random errors. Several parameter estimators have been derived so far to yield parameters with different properties. Some of them were investigated with a Monte Carlo method by Horváth, et al.<sup>4</sup> We can group all parameter estimators into one of 3 classes depending on their derivation:

- Derivation based on geometrical concepts.
- Derivation based on functionals applied to distribution functions.
- Derivation based on test statistics.

#### Geometrical concepts

All these procedures are based on the minimization of vector norms. The most frequently used vector norm is the Euclidean one and least squares estimation may be derived from it. However, it belongs in a more general norm family, in particular  $L_p = \left[ \frac{1}{n} \sum |x_i|^p \right]^{\frac{1}{p}}$ . Minimization of  $L_1$  leads to a linear

programming problem<sup>5,6</sup>. In the case of  $L_2$ , the well-known least squares method can be obtained<sup>7</sup>.  $L_\infty$  plays an important role in the statistics of extremes<sup>8</sup> and in the *minimax*<sup>9</sup> problems.  $L_\infty$  minimization is again a linear programming task. For the work reported here, the following distance measure is defined<sup>6</sup>:  $P = \left\{ \prod [\varepsilon^2 + x_i^2] \right\}^{\frac{1}{2n}}$ , where  $\varepsilon$  is defined by the following recursive formula: 
$$\varepsilon_{(k+1)}^2 = 3 \sum \frac{x_i^2}{(\varepsilon_{(k)}^2 + x_i^2)^2} \bigg/ \sum \frac{1}{(\varepsilon_{(k)}^2 + x_i^2)^2}$$
.  $\varepsilon$  is a type of dispersion of the data like standard deviation. The name of the estimator obtained by minimizing  $P$  is the *most frequent value*<sup>10</sup>. The last norm is derived from the  $L_p$  exchanging summation for the median operator:  $M_p = \left[ \text{med} \{ x_i |^p \} \right]^{\frac{1}{p}}$ . The literature deals only with  $p = 2$ , and the derived procedure is the *least median of squares*<sup>11</sup>.

Functionals applied to distribution functions

In most practical cases, it can be assumed that any estimation depends only on the empirical distribution of the data, i.e., the estimator can be regarded as a functional (assigning a real number to each member of a class of functions) on the empirical distribution function  $F_n$ . However, the estimator  $T$  is usually derived from some theoretical probability distribution  $F$ , denoted by  $T(F)$ .  $T(F_n)$  is a natural nonparametric estimator, but its properties can be deduced by studying the behavior of  $T(G)$  for  $G$  in the neighborhood of  $F$ . This problem concerns the first type of model error, i.e., there is a model to describe the connection between the responses and concentrations (this is the calibration function), including the error distribution of the measurements, but the latter may not be exact. We may assume a distribution function  $F$  instead of  $G$ . The definition of the *qualitative robustness*<sup>13</sup> for estimates follows: *the estimation is robust if the distribution of estimates is a uniformly continuous functional of the mother distribution, i.e., near-lying distributions of estimates belong to near-lying mother distributions* (the Prohorov distance between  $G$  and  $F$  is small). However, the robustness of the estimators will be measured quantitatively in two different ways. The overall sensitivity of an estimator is measured by its breakdown point<sup>9,13,14</sup> (this is often described as *quantitative robustness*), which is roughly the smallest fraction of outliers in the collection that can take the estimator over all

bounds. It describes the overall behavior of an estimator under large perturbations. The influence function<sup>15</sup> (the *infinitesimal* robustness) is defined to measure robustness locally. It describes how the estimator alters under infinitesimal perturbations at some single point. It is seen that a robust estimator can treat the first-mentioned model error if it has a high breakdown point and a bounded, continuous influence function.

Clancey<sup>16</sup> pointed out in his comprehensive study that only 10-15 % of the distributions could be regarded as normal, and 45-50 % of them were symmetrical. That is why we should be interested in several different kinds of error distribution. We now must find procedures that use some assumed distributions to yield estimators. We have to mention the method of moments<sup>17</sup>, though the solution given by it is not always unique. Additionally the higher moments are increasingly more sensitive to experimental error because of the large exponents. For this reason, the method of moments is not robust, and it is therefore not recommended for the estimation of parameters. Frequently used regression procedures can be obtained using the principle of maximum likelihood (ML)<sup>17</sup>. Huber<sup>9</sup> generalized the method of ML by replacing  $-\ln f$  ( $f$  is the identical probability density function of independent observations) by a differentiable function  $\rho$ . The estimators defined by this way are M-estimators and they can be robust. Previously we assumed that we knew the type of the error distribution exactly. However, it frequently occurs that we cannot know it, and the unknown distribution function must be substituted by some given one. If a given distribution of density function  $f(x)$  is substituted by some other distribution of density  $g(x)$ , the relative information or *I-divergence*<sup>18,19,20,10</sup> is defined by the expression

$$I_g(f) = \int_{-\infty}^{\infty} f(x) \log_2 \frac{f(x)}{g(x)} dx. \quad (2)$$

The *I-divergence* can be used in the following practical way. Substitution of the actual but unknown distribution  $f$  by  $g$  of known analytical form minimizes the loss of information (i.e., the *I-divergence*  $I_g(f)$ ). The pair of location ( $T$ ) and scatter ( $S$ ) values of  $g$  will be accepted as characteristics of  $f$ .

Test statistics

Minimizing the well-known  $\chi^2$ -statistic gives the least squares estimator<sup>17</sup>.  $R$ -estimators<sup>9,21</sup> can be obtained by the general scores statistic. Based on the rank correlation tau (i.e., Kendall's tau<sup>22,23</sup>) the following parameter estimation can be derived for the linear relationship  $y = a x + b$ . We can consider the set of  $N$  distinct pairs:  $\hat{b}_{ij} = \frac{y_j - y_i}{x_j - x_i}$ ,  $x_j > x_i$ . The median of ordered  $\hat{b}_{ij}$  can be chosen to estimate  $b$ . For estimation of  $a$  one can calculate the median of ordered  $y_i - b x_i$ . This is the theoretical background of the algorithm has appeared previously<sup>24</sup>. The  $U$ -statistic was introduced by Hoeffding<sup>25</sup> and Siegel<sup>26</sup> robustified it, replacing the overall mean for the nested medians:  $\hat{\theta} = \text{median}^m \{ \theta(x_{\alpha_1}, \dots, x_{\alpha_m}) \}$ . The median operator in turn reduces the number of subscripts by 1 beginning with  $\alpha_m$  and finally reaching  $\alpha_1$ .

MODEL ERROR CAUSED BY DIFFERENCE FROM THE ASSUMED LINEARITY OF THE CALIBRATION FUNCTION

In widespread measuring models, noise is considered to contaminate the true values, for example in Equation 1, and therefore the predicted parameters and the evaluated values are disturbed by errors. Alternatively some of the apparent measurement errors are due to the shortcomings of the applied model (e.g., calibration function). In this case we assume that we can measure precisely, but we are not sure whether the applied model is proper or not. We must choose a model, and the one selected is often very simple to ensure easy manageability. In fact, it is impossible to consider all the relationships practically; we must neglect some of them. The estimation errors, i.e., the distance between measurement and its prediction, therefore exist because of the inappropriateness of the theory describing the system. This error can be interpreted as the difference between the unknown, true model and the applied one, i.e., the deviations are closely related to the fuzziness of the system parameters rather than to observation errors. This problem is no longer a stochastic one, because the system has no random properties, and thus the solution can be obtained by using fuzzy theory. The assumption of linearity is very frequent in analytical chemistry. Hence we present an approach to treat the discrep-

ancy from the assumed linearity following Tanaka, et al.<sup>27</sup> In analytical chemistry we must often use a linear or linearized model for calibration. However, the conditions for this are rarely fulfilled. Thus, we assume linearity, knowing that this can not be quite correct, so this decision will be influenced by human estimation. This problem can be solved by using fuzzy functions  $F(y)$ . In these functions, the parameters are fuzzy numbers<sup>28</sup>, i.e.,  $\tilde{Y} = f(x, \tilde{A})$ . Now  $F(y)$  is the set of all fuzzy subsets on  $\tilde{Y}$ . For the fuzzy linear function  $\tilde{Y} = \tilde{A}_1 x_1 + \dots + \tilde{A}_j x_j + \dots + \tilde{A}_p x_p$  in consequence of the triangular fuzzy number  $\tilde{A}$ , the membership function of  $\tilde{Y}$  is as follows

$$\mu_{\tilde{Y}}(y) = \begin{cases} 1 - \frac{|y - \underline{x}'\underline{\alpha}|}{\underline{c}'|\underline{x}|}, & \underline{x} \neq 0 \\ 1, & \underline{x} = 0, \underline{y} = 0 \\ 0, & \underline{x} = 0, \underline{y} \neq 0 \end{cases} \quad (3)$$

where  $|\underline{x}| = (|x_1|, \dots, |x_n|)$ , and  $\underline{\alpha}$  and  $\underline{c}$  denote the vectors of center values and spreads for all parameters, respectively. The estimated fuzzy linear model should then cover all the data points  $y_i$  with a certain threshold  $H$  chosen by the decision-maker, i.e.,  $\mu_{\tilde{Y}}(y_i) \geq H$  for all  $i=1, 2, \dots, n$  relating to the given non-fuzzy, i.e., precise data  $y_i$ . These inequalities can be rewritten by using the membership function  $\mu_{\tilde{Y}}(y)$ :  $(1 - H)\underline{c}'|\underline{x}| - |y - \underline{x}'\underline{\alpha}| \geq 0$ ,  $\underline{x} \neq 0$ . It seems a natural requirement to choose parameters with which the sum of spreads  $c_j$  will become the smallest; in other words, the solution must be the least vague fuzzy linear function relating to the observations and the well-chosen  $H$ .

It turned out, however, that the fuzzy regression described above was not robust enough. That is why we used an iterative procedure to improve the robustness. Membership values for data  $y_i$  can be calculated with the calculated parameters  $\alpha_j$  and their widths  $c_j$ . Using these membership values, one can construct the fuzzy output data  $(y_i, e_i)$ , where  $e_i$  is a width which is not in inverse ratio to the

membership values of  $y_i$ . It may be surprising, but Fig. 2 in Ref. 29 reveals that the higher membership value of  $y_i$  belongs to the wider  $e_i$ . Thus, in the first step we calculated with the nonfuzzy data (i.e., all  $e_i$  equal 0) and after that with the fuzzy output, obtaining  $e_i$  from the results of the previous step. This iterative procedure is repeated until more than half of the measurements have membership values of  $H$  or more.

Finding the parameters for fuzzy regression problem with fuzzy outputs becomes the following linear programming task<sup>29</sup>:

$$\begin{aligned}
 & \text{Minimize} && s = \sum_{j=1}^p c_j \\
 & \text{Subject to} && (1-H) \sum_{j=1}^p c_j |x_{ij}| + \underline{x}^t \underline{\alpha} \geq y_i + (1-H)e_i, \\
 & && (H-1) \sum_{j=1}^p c_j |x_{ij}| + \underline{x}^t \underline{\alpha} \leq y_i + (H-1)e_i, \\
 & && c_i \geq 0, \\
 & && \text{for all } i = 1, 2, \dots, n.
 \end{aligned}
 \tag{4}$$

EXPERIMENTAL RESULTS

Real calibration data from Inductively Coupled Plasma Atomic Emission Spectrometry (ICP-AES), obtained for subsurface- and drinking water environmental analysis, were investigated by using the following methods: Least Sum of Squares (LS), Least Sum of Absolute Residuals (LSA)<sup>6</sup>, Least Maximum Absolute Residuals (LMA)<sup>6</sup>, Iteratively Reweighted Least Sum of Squares with Tuning Constants 6 and 9 (IRLS6 and IRLS9)<sup>30</sup>, Most Frequent Value (MFV)<sup>10</sup>, Single Median (SM)<sup>23,24</sup>, Repeated Median (RM)<sup>26</sup>, Least Median of Squares (LMS)<sup>11,12</sup>. Data are listed in Table 1 and the results are given in Table 2.

Overall, it may be stated that these data were selected from several calibration data sets to demonstrate the effects of outliers and the usefulness of robust procedures. Outliers existed because of occasional errors of unknown origin in instrumentation or sample preparation.

It is well-known that an ICP spectrometer can produce linear readings over six orders of magnitude in concentration. Because of this, its data represents the best example of the first type of model

TABLE 1. Calibration Data Measured by ICP-AES.

Concentration (ppm)	Signals for Elements (arbitrary units)					
	Mo	Cr	Co	Pb	Ni (221.6 nm)	Ni (231.6 nm)
0	8.19	-23.4	-1.83	13.46	6.4	28.33
0	16.05	-19.4	-2.45	6.4	7.47	30.56
0.25	171.9	210.9	261.3	112	223.7	220.8
0.25	180.6	213.7	260.1	119.2	215.6	218.6
0.5	406	420.4	430.8	217	437.9	410.2
0.5	414.5	423.2	431.4	207.7	430.7	407.9
1	810.7	843.2	860.3	419.6	897.3	828.3
1	818.2	840.6	859.6	428.7	886.8	826.1

TABLE 2. Estimated Parameters Obtained by the Methods Mentioned in the Text for the Data in Table 1.

Methods	Mo		Cr		Co		Pb		Ni*		Ni**	
	a <sub>0</sub>	a <sub>1</sub>	a <sub>0</sub>	a <sub>1</sub>	a <sub>0</sub>	a <sub>1</sub>	a <sub>0</sub>	a <sub>1</sub>	a <sub>0</sub>	a <sub>1</sub>	a <sub>0</sub>	a <sub>1</sub>
LS	-3.07	814.5	-11.9	858.5	17.2	846.1	9.84	413.0	0.47	886.3	21.9	798.7
LSA	8.19	802.5	-2.4	845.6	1.3	859.0	13.5	407.1	2.67	884.1	19.0	807.1
LMA	-6.29	802.1	-12.8	864.0	21.7	862.1	8.7	412.7	-3.37	889.8	19.8	797.7
IRLS6	11.1	802.3	-11.7	858.0	-1.45	862.1	10.0	412.9	0.85	885.8	22.2	799.1
IRLS9	-1.88	813.6	-11.8	858.3	15.7	847.2	9.91	412.9	0.63	886.1	22.0	798.8
MFV	9.36	803.9	1.31	840.8	-1.96	862.1	10.3	412.1	6.4	880.4	22.9	801.3
SM	5.25	808.4	-9.15	854.6	0.39	860.2	9.65	412.9	5.1	880.0	26.7	782.7
RM	2.01	813.5	-1.55	844.5	2.25	857.7	10.4	410.8	6.2	870.92	28.9	765.2
LMS	12.1	802.5	2.4	839.6	-1.34	862.0	11.8	407.1	8.98	848.6	29.4	761.1

\* measured at wavelength 221.6 nm

\*\* measured at wavelength 231.6 nm

error, where the linear relationship is adequate, but the assumed error distribution can fail, and some outliers can occur within measurements. For the data shown, the best results were given by LMS. MFV also yielded appropriate results for the measurements of Mo, Cr, Pb, Co and Ni (221.6 nm). The calibration lines calculated by RM was acceptable for Cr, Pb and Ni measured at both wavelengths, and by IRLS6 for Mo and Co. SM was good for Pb and Ni (231.6 nm) and LSA for only Mo. IRLS9 and LS gave nearly the same results. LMA gave very biased parameters, in fact it was the most sensitive to the outliers.



TABLE 3. Results of Fuzzy Linear Regression and its Computation by Our Modification.

		Data Set <sup>3</sup>		
		No. 2.	No. 5.	No. 8.
1. iteration	$a_0 \pm c_0$	$0.0 \pm 0.0$	$0.019 \pm 0.000$	$0.0 \pm 0.0$
	$a_1 \pm c_1$	$0.00447 \pm 0.00043$	$0.00400 \pm 0.00010$	$0.0100 \pm 0.0014$
	$\mu_{Y_1}$	1.000	0.100	1.000
	$\mu_{Y_2}$	0.373	0.100	0.308
	$\mu_{Y_3}$	0.100	0.614	0.100
	$\mu_{Y_4}$	0.100	0.100	0.481
	$\mu_{Y_5}$	0.197	0.923**	0.573
	$\mu_{Y_6}$			0.697
2. iteration	$a_0 \pm c_0$	$0.0 \pm 0.0$		$0.0 \pm 0.0$
	$a_1 \pm c_1$	$0.00408 \pm 0.00776$		$0.0870 \pm 0.0016$
	$\mu_{Y_1}$	1.000		1.000
	$\mu_{Y_2}$	0.985		0.861
	$\mu_{Y_3}$	0.901		0.843
	$\mu_{Y_4}$	1.000		0.876
	$\mu_{Y_5}$	0.995*		0.884
	$\mu_{Y_6}$			0.947
3. iteration	$a_0 \pm c_0$			$0.0 \pm 0.0$
	$a_1 \pm c_1$			$0.0104 \pm 0.0106$
	$\mu_{Y_1}$			1.000
	$\mu_{Y_2}$			0.962
	$\mu_{Y_3}$			0.944
	$\mu_{Y_4}$			0.977
	$\mu_{Y_5}$			0.985
	$\mu_{Y_6}$			0.951
			$0.899^*$	

\* The stop criterion was fulfilled when more than half of the measurements had a membership value of 0.9 (=  $H$ ) or more.

\*\* The stop criterion was fulfilled when more than half of the measurements had a membership value of 0.1 (=  $H$ ) or more.

Table 4. Properties of the Parameter Estimators Investigated

Methods	Derivations	Breakdown points	Robustness	Recommendations
LMA	$\min L_{\infty}$	0 %	not robust at all	not recommended at all
LS	$\min L_2$	0 %	not robust at all	recommended only for large data
LSA	$\min L_1$	0 %	sensitive to leverage points*	(30-50 measurements) recommended only with good experimental design
IRLS9	M-estimator	20 %	slightly robust	recommended only for large data
IRLS6	M-estimator	25 %	robust	} (15-30 measurements) recommended to use in calibration even for small data (8-15 measurements)
MFV	M-estimator	25 %	robust	
SM	based on rank correlation	30 %	robust	
RM	based on modified U-statistic	50 %	very robust	
LMS	$\min M_2$	50 %	very robust	

\* highly influential data points caused by non balanced design

Some calibration data from atomic absorption spectrometry was recently investigated by Hu, et al.<sup>3</sup> We carried out fuzzy linear calibration for three selected data sets investigated by them: those for data sets 2, 5 and 8. These were of interest because they were measured with gross outlier, no outliers and with model error, respectively. The results are shown in Table 3. The modified fuzzy linear regression method worked well in all three of these examples.

## DISCUSSION

This article intended to demonstrate the problem of two types of model error, which in our opinion is a central point of the parameter estimation and therefore the chemical calibration.

Table 4 shows the properties of the parameter estimators reported here to make clear their effective use in calibration based on theoretical and practical considerations.

It should be noted that the first type of model error exists because of the lack of knowledge about the exact distribution of the measurement error. Nevertheless, not even complete knowledge can help in the case of the small calibration data sets<sup>12</sup> often used in practice. When the number of measure-

ments is small, the empirical distribution can differ completely from that theoretically expected. It may even be asymmetric, though symmetry may be assumed for large data. Thus, outliers can occur much more frequently than expected. If an estimator has a bounded, continuous influence function, it is robust. However, we must argue with the view<sup>31</sup> that one should choose the biweight procedure<sup>30</sup> instead of the median, just because the former has better properties in its influence function than the latter has for special, artificially compiled data. In fact, a robust procedure needs a high breakdown point too, because near this point the result may be rather biased (see Fig. 2 on page 42 in Ref. 32); it should not be forgotten that the bias tends to infinity on collapsing. In addition, the influence function depends on the actual error distribution too, while the breakdown points are distribution-free for most of the cases. Outliers can occur more than one at a time, it is true that their ratio is not less than 10-20 %, in general. All the same, it is safer to use an estimator which has a higher breakdown point.

If one has more information about the type of the error distribution for the instrument used, then it is possible to use functionals applied to them. The principle of maximum likelihood and its generalization can give either nonrobust or robust estimators. However, it can often occur that the estimated scale is biased and/or nonrobust, making it more difficult to obtain a proper location. The minimization of *I*-divergence can yield robust scales. Additionally, its base concept, that the practical distribution is not known exactly, and is instead substituted by a theoretical one, is closer to practice in chemical calibration. Unfortunately, there is no possibility to create confidence interval for the estimated parameters. However, this problem is general in robust statistics, a result of the difficulty of describing the distribution of the outliers. Work addressing this question will appear soon<sup>33</sup>. Walters and Rizzuto<sup>34</sup> investigated nonrobust (inverse and classical) estimators to compare them based on the confidence interval.

In deciding when to use robust procedures a quality coefficient (*QC*)<sup>35</sup> can help. This quality coefficient is defined as  $QC = 100 \cdot \sqrt{\frac{\sum_{i=1}^n \left( \frac{y_i - \hat{y}_i}{y_i} \right)^2}{(n-1)}}$ , where  $y_i$  is the measurement,  $\hat{y}_i$  is the prediction by LS line and  $n$  is the number of data points, without the zero points. A critical value

is given in advance to considering the precision of the analytical method. If the  $QC$  value is larger than the critical value, the hypothesis that there are no outliers will be rejected and robust procedures will be used. For the ICP data a threshold value of 10 % was used. The  $QC$  values were found 69.2 %, 23.7 %, 496.4 %, 22.9 %, 49.8 % and 13.8 % for measurements according to the elements Mo, Cr, Co, Pb and Ni (221.6 nm and 231.6 nm), respectively. In the case of Co the  $QC$  value was rather high. It turned out, that there were gross outliers at concentration 0.25 ppm and the LS line was rather far from the signal measured at concentration 0 ppm.

The second type of model error is a bit more difficult. We gave a solution when the linearity was assumed followed by concepts of Tanaka et al., and the robustness of the method was improved. The second type of model error generally arises when an approximately linear segment of nonlinear response function is used. This is the case in the given examples of AAS measurements, and we illustrated how to treat the failing linearity. Recently a pleasant solution appeared for diagnosing the linearity of calibration lines by Vankeerberghen, et al.<sup>36</sup> We think, however, that the fuzzy linear regression procedure with our modification needs no limitation of number of data points and it can also work in multivariate case. More than one iteration cycle means that the assumption of linearity should be rejected; however, the obtained parameters can be used for evaluation because of the robust property of the estimation.

We should also mention that real problem can emerge in automatic analyses. In this case there is not always the chance to control the calibration graphs visually, one by one, either to recognize outliers or to check linearity. Algorithms must automatically treat failures caused by either of model errors. The diagnostic can be carried out as described before. Of course, after a suitable period, the analyst must check whether discrepancies occurred in calibration and he or she must correct the analytical conditions to try to avoid outliers next time.

#### REFERENCES

1. M. Otto and H. Bandemer, *Chemom. Int. Lab. Syst.*, 1, 71-78 (1986).
2. R. Rajkó, I. Horváth and P. Huhn, *Magy. Kém. Foly.*, 95, 323-326 (1989).

3. Y. Hu, J. Smeyers-Verbeke and D.L. Massart, *Chemom. Int. Lab. Syst.*, **8**, 143-155 (1990).
4. I. Horváth, R. Rajkó and P. Huhn, *Magy. Kém. Foly.*, **95**, 327-335 (1989).
5. S.P. Ellis and S. Morgenthaller, *J. Amer. Stat. Assoc.*, **87**, 143-148 (1992).
6. R.L. Sileken, Jr. and H.O. Hartley, *J. Amer. Stat. Assoc.*, **68**, 639-641 (1973).
7. S.R. Searle, *Matrix Algebra Useful for Statistics*, Wiley, New York, 1982.
8. E.J. Gumbel, *Statistics of Extremes*, Columbia University Press, 1958.
9. P.J. Huber, *Robust Statistics*, Wiley, New York, 1981.
10. F. Steiner, *Geoph. Trans.*, **34**, 139-260 (1988).
11. P.J. Rousseeuw, *J. Amer. Stat. Assoc.*, **79**, 871-880 (1984).
12. P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection*, Wiley, New York, 1987.
13. F.R. Hampel, *Ann. Math. Stat.*, **42**, 1887-1896 (1971).
14. D.L. Donoho and P.J. Huber, The notion of breakdown point, in P.J. Bickel, K.A. Doksum and J.L. Hodges Jr. (Editors), *A Festschrift for Erich L. Lehmann*, Wadsworth, Belmont, California, 1983, pp. 157-184.
15. F.R. Hampel, *J. Amer. Stat. Assoc.*, **69**, 383-393 (1974).
16. V.J. Clancey, *Nature*, **159**, 339-340 (1947).
17. I. Vincze, *Mathematische Statistik mit industriellen Anwendungen*, Akadémiai Kiadó, Budapest, 1971.
18. S. Kullback, *Information Theory and Statistics*, Wiley, New York, 1959.
19. A. Perez, *Kybernetika*, **3**, 1-21 (1967).
20. J.E. Shore and R.W. Johnson, *IEEE Trans. Inf. Theory*, *IT-26*, 26-37 (1980). [Correction *IT-29*, 942-943 (1983)].
21. J. Jureckova, *Ann. Math. Stat.*, **42**, 1328-1338 (1971).
22. M.G. Kendall, *Rank Correlation Methods*, Griffin, London, 1955.
23. P.K. Sen, *J. Amer. Stat. Assoc.*, **63**, 1379-1389 (1968).
24. M. Feinberg, *Trends Anal. Chem.*, **4**, 189-190 (1985).
25. W. Hoeffding, *Ann. Math. Stat.*, **19**, 293-325 (1948).
26. A.F. Siegel, *Biometrika*, **69**, 242-244 (1982).
27. H. Tanaka, S. Uejima and K. Asai, Fuzzy linear regression model, *International Congress on Applied Systems Research and Cybernetics*, Acapulco, Mexico, 1980.
28. D. Dubois and H. Prade, *Fuzzy Sets and Systems. Theory and Applications*, Academic Press, New York, 1980.
29. H. Tanaka, S. Uejima and K. Asai, *IEEE Trans. Syst., Man Cyber.*, *SMC-12*, 903-907 (1982).
30. G.R. Phillips and E.M. Eyring, *Anal. Chem.*, **55**, 1134-1138 (1983).

31. P. Vankeerberghen, C. Vandebosch, J. Smeyers-Verbeke and D.L. Massart, *Chemom. Int. Lab. Syst.*, *12*, 3-13 (1991).
32. F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel, *Robust Statistics: the Approach Based on Influence Functions*, Wiley, New York, 1986.
33. R. Rajkó, I. Horváth, S. Veres and A. Csikkel-Szolnoki, in preparation.
34. F.H. Walters and G.T. Rizzuto, *Anal. Lett.*, *21(11)*, 2069-2076 (1988).
35. J. Knecht and G. Stork, *Fresenius' Z. Anal. Chem.*, *270*, 97-99 (1974).
36. P. Vankeerberghen, J. Smeyers-Verbeke, A Thielemans and D.L. Massart, *Analisis*, *20*, 103-109 (1992).

Received June 14, 1993

Accepted August 27, 1993