Talanta, Vol. 40, No. 2, pp. 279–285, 1993
Printed in Great Britain.

0039-9140/93 $6.00 + 0.00

MULTIPARAMETRIC CURVE FITTING XV

STATISTICAL ANALYSIS AND GOODNESS-OF-FIT TEST BY THE LEAST-SQUARES ALGORITHM MINOPT

JIŘÍ MILITKÝ
Department of Textile Materials, Technical University, CS-461 17 Liberec, Czech Republic

MILAN MELOUN
Department of Analytical Chemistry, University of Chemical Technology, CS-532 10 Pardubice, Czech Republic

(*Received* 9 *December* 1991. *Revised* 29 *June* 1992. *Accepted* 29 *June* 1992)

Summary—Estimation of nonlinear regression quality leads to examination of quality of parameter estimates, a degree of fit, a prediction ability of model proposed and quality of experimental data. Statistical analysis serves for computation of confidence intervals of parameters and confidence bands, the bias of parameters and bias of residuals. Goodness-of-fit test examines classical residuals using various diagnostics and identifies influential points. Mentioned topics of nonlinear model building and testing contained in MINOPT program from CHEMSTAT package are illustrated.

Practical applicability of regression algorithms and program packages for non-linear regression can be deduced from an ability to reach a minimum of a sum of squared residuals and from a quality and amount of statistical information. Structural classification of regression programs in blocks already introduced in ABLET programs of solution equilibria[1,2] and instrumental methods of analytical and physical chemistry concerns blocks INPUT, RESIDUAL SUM OF SQUARES, MINIMIZATION, STATISTICAL ANALYSIS, GOODNESS-OF-FIT TEST, DATA SIMULATION, *etc.* was also used here.

While a previous paper[3] of this series describes RESIDUAL SUM OF SQUARES and MINIMIZATION blocks, this paper brings a description of two other blocks of MINOPT structure, *i.e.*, STATISTICAL ANALYSIS and GOODNESS-OF-FIT TEST. Procedure of regression model testing[4] is illustrated.

THEORY

Statistical analysis block

Statistical analysis in nonlinear regression depends on an actual model used, measurement errors and a criterion function. Let us concentrate here on the method of maximum likelihood when the searched estimates $\hat{b}$ maximize the logarithm of the likelihood function $l(\beta) = \ln L(\beta)$. If for the additive model of measurement (*cf.* Ref. 3) the independent errors ϵ have the probability density function $p(\epsilon)$ then likelihood function $L(b)$ is defined as

$$L(b) = \prod_{i=1}^{n} p(y_i - f(x_i; b) \qquad (1)$$

In construction of confidence intervals of parameters β or in statistical hypotheses testing the linearization, Lagrange multipliers and likelihood ratio methods may be used.[5]

The least-squares (LS) method is the best case of an additive model of measurement and independent normally distributed measurement errors having constant variance. Gallant[5] shows that the least squares estimator $\hat{b}$ of true value of parameters β in the regression model has asymptotically m-dimensional normal distribution

$$\hat{b} \approx N[\beta, \sigma^2(\mathbf{J}^T\mathbf{J})^{-1}] \qquad (2)$$

Here σ^2 is error variance and $\mathbf{J}$ is the Jacobi matrix (definition, *cf.* ref. 3). The asymptotic normality of estimates $\hat{b}$ determined by the least-squares method does not require a normality of errors ϵ, Ref. 5.

For real experimental data the estimates $\hat{b}$ and other statistical characteristics are biased and therefore application of equation (2) is

limited. Statistical analysis of nonlinear regression models by the least-squares methods then depends on a magnitude of bias which describes a degree of nonlinearity of regression model.

Covariance matrix of parameter estimates. From equation (2) it follows that the asymptotic covariance matrix of estimate $\hat{b}$ is expressed by the relation

$$D(\hat{b}) = \sigma^2(\mathbf{J}^T\mathbf{J})^{-1} \tag{3}$$

where s^2 is estimator of σ^2. There exist many more accurate expressions,[6] but for practical calculations the asymptotic formula [equation (3)] is quite acceptable.

On the base of knowledge of a covariance matrix $D(\hat{b})$ either the variance of individual parameters $D(\hat{b}_j)$ or the correlation coefficients r_{ij} between estimates $\hat{b}_i$ and $\hat{b}_j$ may be estimated.

Bias of parameter estimates. The bias is given by

$$h = E(\hat{b} - \beta^*) \tag{4}$$

For the sake of simplicity we use an expression of parameter bias in the form[7]

$$h = (\mathbf{J}^T\mathbf{J})^{-1}\mathbf{J}^T d \tag{5}$$

where d is the $(n \times 1)$ vector with components

$$d_i = \frac{-\sigma^2 \operatorname{tr}[(\mathbf{J}^T\mathbf{J})^{-1}\mathbf{W}_i]}{2} \tag{6}$$

where tr[·] denotes a trace of matrix and $\mathbf{W}_i$ is the matrix of second derivatives of model function in the i-th point. The vector d is an expected value of difference between the linear and quadratic approximation of a model function.

Similarly the bias of residuals

$$\hat{e}_i = y_i - \mathrm{f}(x_i; \hat{b}) \tag{7}$$

can be defined. When $E(\epsilon) = 0$, the bias of residuals is equal to their mean value $E(\hat{e})$. The mean value of residuals vector

$$E = E(\hat{e}) \tag{8}$$

can be rewritten[7] as

$$E = (\mathbf{E} - \mathbf{P})\, d \tag{9}$$

where $\mathbf{P} = \mathbf{J}(\mathbf{J}^T\mathbf{J})^{-1}\mathbf{J}^T$ is the projection matrix and $\mathbf{E}$ is the unit matrix of order n.

For practical calculation the relative bias of parameter estimates is often used

$$h_{Rj} = \frac{h_j}{\hat{b}_j} 100 \quad [\%] \tag{10}$$

The bias of estimates is considered significant if $h_{Rj} > 1\%$ holds.[7] For such biased estimates the statistical analysis based on linearization of regression model cannot be correctly used.

For expressing the total bias of parameter estimates Box[8] proposed the scalar characteristic

$$\hat{M} = \frac{h^T(\mathbf{J}^T\mathbf{J})h}{m\hat{\sigma}^2} \tag{11}$$

The bias of parameters may be affected by a reparametrization.[9]

Interval estimates of parameters. Points estimates $\hat{b}$ of regression parameters β are, in the statistical view, worthless as they do not mention intervals in which a true value β may be expected. The estimates $\hat{b}$ are random quantities estimated on base of sample, $\{y_i, x_i\}$, $i = 1, \ldots, n$.

In nonlinear regression models for a construction of confidence regions and intervals a linearization is often used for which confidence regions are elliptic. However, a linearization is useful only in cases when a model is not strongly nonlinear and nonlinearity measures, for example, the parameter bias, are small. The more accurate confidence region calculated on the base of Lagrange multipliers or the likelihood ratio can also be constructed. They are generally non-elliptic and are not continuous.

For asymptotic normality of maximum likelihood estimates $\hat{b}$ it follows that the quadratic form

$$Q = (\beta - \hat{b})^T D(\hat{b})^{-1}(\beta - b) \tag{12}$$

has χ^2(m) distribution. The corresponding 100(1–α)% confidence region of parameters β forms a m-dimensional ellipsoid with boundaries expressed by

$$(\beta^* - \hat{b})^T D(\hat{b})^{-1}(\beta^* - \hat{b}) = \chi^2_{1-\alpha}(m) \tag{13}$$

where $\chi^2_{1-\alpha}(m)$ is the $100(1-\alpha)$%th quantile of $\chi^2(m)$ with m degrees of freedom. The center of this ellipsoid is in the point $\hat{b}$.

For the least-squares method the application of equation (13) leads to definition of confidence ellipsoid having the boundary

$$\Delta b^T(\mathbf{J}^T\mathbf{J})^{-1}\,\Delta b = m\hat{\sigma}^2 F_{1-\alpha}(m, n-m) \tag{14}$$

where $\Delta b = \beta - \hat{b}$ and $F_{1-\alpha}(m, n-m)$ is quantile of Fisher–Snedecor distribution.

When a bias of parameters h is calculated, instead of Δb the correction $\Delta b_0 = \hat{b} - h - \beta$ may be used.

For expressing a geometry of confidence ellipsoids the decomposition of the matrix $(\mathbf{J}^T\mathbf{J})^{-1}$ to eigenvalues L_i and eigenvectors Z_i may be introduced

$$(\mathbf{J}^T\mathbf{J})^{-1} = \mathbf{ZLZ}^T \tag{15}$$

where $\mathbf{Z}$ is a matrix containing eigenvectors in columns and diagonal matrix $\mathbf{L}$ contains eigenvalues $L_1 \geqslant L_2 \geqslant \cdots \geqslant L_m$ on a diagonal. Using this decomposition the new orthogonal set of coordinates $y = \mathbf{Z}\,\Delta b$ can be defined. This set has an important property that the axes of confidence ellipsoid are identical with the axes of the coordinate system. Introducing notation

$$p^2 = m\hat{\sigma}^2 F_{1-\alpha}(m, n-m) \tag{16}$$

the confidence ellipsoid can be expressed by simple formula

$$\sum_{i=1}^{m} \frac{Y_i^2}{L_i} = p^2 \tag{17}$$

The lengths of half-axes of the ellipsoid are equal to $p\sqrt{L_i}$. For a projection Δ_{jk} of the j-th half-axis into the axis of parameter β_k it holds that

$$\Delta_{jk} = p|Z_{kj}\sqrt{L_j}| \tag{18}$$

where Z_{kj} is the k-th elements of the vector Z_j which is the j-th column of matrix $\mathbf{Z}$.

When dimension of a parameter vector is $m > 2$, a partial confidence ellipsoid can be constructed.[5]

For building the confidence region the Lagrange multipliers or a likelihood ratio may also be used. For example, from properties of likelihood ratio the bound of $100(1-\alpha)\%$ confidence region can be defined by relation

$$2[\ln L(\hat{b}) - \ln L(\beta)] = \chi^2_{1-\alpha}(m) \tag{19}$$

For a least-squares criterion the relation (19) leads to the relation

$$U(\beta) - U(\hat{\beta}) = m\hat{\sigma}^2 F_{1-\alpha}(m, n-m) \tag{20}$$

The confidence region defined by this equation is not generally elliptical or continuous.

With the use of equation (13) the $100(1-\alpha)\%$th confidence interval of parameter β_j in the form

$$\hat{b}_j - \hat{\sigma}\sqrt{C_{jj}}\, t_{1-\alpha/2}(n-m) \leqslant \beta_j \leqslant \hat{b}_j + \hat{\sigma}\sqrt{C_{jj}}\, t_{1-\alpha/2}(n-m) \tag{21}$$

is direct analogy of confidence intervals of the parameters of linear models. An influence of other parameters is neglected. When all diagonal-off elements of the matrix $\mathbf{C} = (\mathbf{J}^T\mathbf{J})^{-1}$ are zero the relation (21) may be used. However, elements of the vector $\hat{b}$ are often mutually correlated so that intervals of equation (21) are under-estimated.

More suitable determination of the confidence interval of parameter β_k is based on the maximal length Δ_k of a projection Δ_{kj} into the parameter axis β_k. The confidence interval of a parameter β_k is then estimated by

$$\hat{b}_k - \Delta_k \leqslant \beta_k \leqslant \hat{b}_k + \Delta_k \tag{22}$$

Instead of projections it is simpler to search directly coordinates of extreme points on the confidence ellipsoid in directions of individual parameter axes.[2] The corresponding confidence interval of a parameter β_k is defined by inequality

$$\hat{b}_k - p\sqrt{C_{kk}} \leqslant \beta_k \leqslant \hat{b}_k + p\sqrt{C_{kk}} \tag{23}$$

holds. For $m = 1$ all confidence intervals[21-23] are identical. Increasing the number of regression parameters m the confidence intervals (22) and (23) are broader than those of (21). All confidence intervals are symmetrical. Using linearization the confidence intervals of prediction $f(x^*; \hat{b})$ and confidence bands can be simply derived.[4] The more accurate confidence bands may be constructed with the use of convenient reparametrization.[9]

Goodness-of-fit tests block

In many regression programs the statistical analysis of residuals represents the main diagnostic tool and a resolution criterion in a search of the "best" model when more than one are possible or proposed. The goodness-of-fit test (which is also called the fitness test) analyses the residual set and examines statistical characteristics.

To application of statistical analysis of classical residuals $\hat{e}$, it should be critically noted that the diagnostic use of classical residuals is not rigorous but of a rather approximate character. The classical residuals do not exhibit a zero mean, they are biased and they are a combination of errors ϵ. Moreover, they are dependent on true values of parameters β which are unknown.

Statistical analysis of classical residuals. Classical residuals are defined as the differences $\hat{e}$ between observation y_i and prediction $\hat{y}_i = f(x_i; \hat{b})$ by equation (7). Graphical and analytical examining residuals check the quality of a

nonlinear model.[4] The following plots are often used in nonlinear models examination:

(1) The overall diagram gives a first view of residuals. If the model is correct these residuals should resemble observations from a normal distribution with zero mean.

(2) Plot type I (also called the index plot) is a scatter plot of residuals $\hat{e}_i$ against an index i in the time order as occurred.

(3) Plot type II (also called the plot against an independent variable) is a scatter plot of residuals $\hat{e}_i$ against the independent variable x_j, $j = 1, \ldots, m$.

(4) Plot type III (also called the plot against a prediction) is a scatter plot of residuals against the prediction $\hat{y}_i$.

The following statistics are used in nonlinear models examination:

(1) The arithmetic mean of residuals known as the estimate of residuals bias, $E(\hat{e})$, should be equal to zero;

(2) The residual variance is calculated from the residual sum of squares

$$\hat{\sigma}^2 = U(\hat{b})/(n-m) \tag{24}$$

The square-root of a residual variance known as an estimate of the residual standard deviation, $s(\hat{e})$, should be of the same magnitude as the (instrumental) error $s_{\text{inst}}(y)$, of dependent variable (observation, measured quantity y), *i.e.*, $s(\hat{e}) \approx s_{\text{inst}}(y)$;

(3) The determination coefficient D^2 is computed from the relation

$$D^2 = 1 - \frac{U(\hat{b})}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{25}$$

where $\bar{y} = 1/n \, \Sigma_{i=1}^{n} y_i$ is the arithmetic mean of response variable. The determination coefficient is for linear models equal to the square of the multiple correlation coefficient. When the determination coefficient is multiplied by 100% we receive the regression rabat in percents, $100D^2$[%]. Determination coefficient D^2 is an increasing function of a number of parameters, therefore, it is not convenient to use as a resolution diagnostic for search of models of different numbers of parameters.

(4) To distinguish between various models proposed the Akaike information criterion AIC is more suitable to apply being defined by the relation

$$\text{AIC} = -\ln L(\hat{b}) + 2m \tag{26}$$

where $L(\hat{b})$ is the likelihood function. The "best" model is considered to be a model for which this criterion reaches a minimal value. Using the least-squares criterion the AIC may be expressed

$$\text{AIC} = n \ln\left[\frac{U(\hat{b})}{n}\right] + 2m \tag{27}$$

(5) The prediction ability of a model proposed may be examined by the mean quadratic error of prediction MEP being defined by the relation

$$\text{MEP} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \text{f}(x_i; \hat{b}_{(i)})^2 \tag{28}$$

The symbol $\hat{b}_{(i)}$ denotes the estimator of parameter β computed without the point (x_i, y_i). Here instead of the parameter estimate $\hat{b}_{(i)}$, the one-step approximation $\hat{b}^1_{(i)}$ defined by following equation (29) may be used. Lower values of MEP criterion give better prediction ability of the model proposed.

Identification of influential points. Influential points can strongly affect some regression characteristics. The points affecting prediction $\hat{y}_i$, for example, may not be influential from the point of view of parameter variance. The degree of influence of individual points should be classified regarding which characteristics are affected.[4] While for linear models all the characteristics for identification of influential points are a function of residuals $\hat{e}$ and diagonal elements P_{ii} of the projection matrix $\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. For nonlinear regression models the parameter estimates and residuals cannot be expressed so simply as a linear combination of experimental data. When the Taylor type linearization of original nonlinear model is used, all methods of identification of influential points in linear models can be used here. Then in the nonlinear case the matrix $\mathbf{J}$ has the same role as $\mathbf{X}$ in the linear case. For a one-step approximation of the parameter estimate $\hat{b}_{(i)}$ computed without point (x_i, y_i) is valid

$$\hat{b}^1_{(i)} = \hat{b} - (\mathbf{J}^T\mathbf{J})^{-1} J_i \frac{\hat{e}_i}{1 - P_{ii}} \tag{29}$$

Here P_{ii} are elements of a projection matrix $\mathbf{P} = \mathbf{J}(\mathbf{J}^T\mathbf{J})^{-1}\mathbf{J}^T$. With use of equation (29) the variance estimate $\hat{s}^2_{(ii)}$ when leaving out the ith point is defined by

$$\hat{s}^2_{(i)} = \frac{U(\hat{b}) - \dfrac{\hat{e}_i^2}{1 - P_{ii}}}{n - m - 1} \tag{30}$$

Some characteristics of influential points based on linearization and used in program MINOPT are in Table 1. Interpretation of these characteristics may be found in Ref. 4.

To express an influence of individual points on parameter estimates the quadratic expansion of a regression model may be used, too.[10]

Nonlinear measure of an influence of the ith point on the parameter estimates is also represented by the likelihood distance

$$LD_i = 2[\ln L(\hat{b}) - \ln L(\hat{b})_{(i)})] \qquad (31)$$

In case of the least-squares the likelihood distance is expressed by

$$LD_i = n \ln\left[\frac{U(\hat{b}_{(i)})}{U(\hat{b})}\right] \qquad (32)$$

In both equations (31) and (32) the estimate $b_{(i)}$ is calculated by a nonlinear regression when the ith point is left out or the one-step approximation $\hat{b}^1_{(i)}$ of the parameter estimates is used. When inequality $LD_i > \chi^2_{1-\alpha}(2)$ holds, the ith point is strongly influential. The significance level α is usually chosen to be equal to 0.05.

Procedure of nonlinear model testing

A quality of nonlinear model proposed is examined using following criteria.

Quality of parameter estimates. Quality of parameter estimates $\hat{b}_j$ is considered according to their confidence intervals Δ_j, equation (18) and (21) and Δ_{Rj}, equations (22) and (23) or according to their standard deviations $s(\hat{b}_j)$, equation (3), the absolute bias h_j, equation (5) and the relative bias h_{Rj}, equation (10). Often an empirical rule of thumb is used: the parameter β_j is considered to be significant when its estimate $\hat{b}_j$ is greater than its 2 standard deviations, $2s(\hat{b}_j) < |\hat{b}_j|$. High values of parameter standard deviation $s(\hat{b}_j)$ is caused by termination of a minimization process before reaching minimum. Therefore, also inaccuracy of calculation of matrix **J** appears or a high nonlinearity of

Table 1. Three characteristics of influential points based on linearization. Critical level is the value of characteristic exceed this level, the corresponding point is denoted as highly outlying

Name	Form	Critical level
Cook distance D_i	$\frac{(\hat{b}-\hat{b}_i)^T \mathbf{J}^T \mathbf{J}(\hat{b}-\hat{b}_i)}{m\hat{\sigma}^2}$	1
DF_i	$\frac{(\hat{y}_i - \hat{y}_{(i)i})^2}{\hat{s}^2_{(i)} P_{ii}}$	$2\sqrt{m/n}$
Jackknife residual $\hat{e}_{Ji}$	$\frac{\hat{e}_i}{\hat{s}_{(i)}\sqrt{1-P_{ii}}}$	3

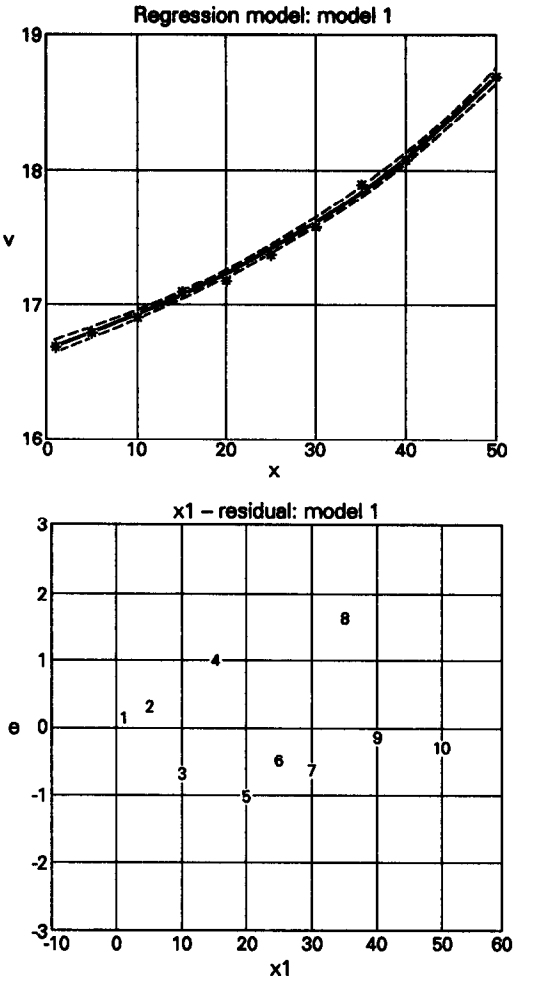

Fig. 1. Non-linear regression of data for Model I: (a) a curve-fitting, and (b) a scatter plot of type II of classical residuals.

regression model exists. The test of statistical significance of each parameter β_j, the null hypothesis $\beta_j = 0$ *vs.* the alternative one $\beta_j \neq 0$, is carried out.

Inter-dependence between parameters. Matrix of paired correlation coefficients of parameters, r_{ij}, expresses a measure of correlation or inter-dependence between two parameters β_i and β_j. If r_{ij} in absolute value is close to one, two parameters β_i and β_j are linearly dependent.

Quality of achieved model fitness. Agreement of proposed model with experimental data is examined by (i) the statistical analysis, and (ii) the goodness-of-fit test.

The statistical analysis of nonlinear regression contains following characteristics: the residual sum of squares $U(\hat{b})$, the regression rabat in percents 100 D^2[%], equation (25), the mean quadratic error of prediction MEP, equation (28), the Akaike information criterion AIC,

Table 2. Illustration of shortened output of MINOPT analysis of $\{x_i, y_i\}$ data for Model I

Quality of parameter estimates
Point and interval estimates of parameters

Parameter	Point estimate	Standard deviation	Absolute bias	Relative bias	Half-length of confidence interval	
β_j	$\hat{b}_j$	$s(\hat{b}_j)$	h_j	$h_{R,j}$, [%]	Δj	$\Delta_{R,j}$
β_1	1.5673E + 01	1.7261E − 01	−0.0161	−0.1028	±0.6232	±0.6233
β_2	9.9925E − 01	1.5625E − 01	0.0160	1.5977	±0.5637	±0.5642
β_3	2.2222E − 02	2.1017E − 03	3.9E − 06	0.0176	±0.0075	±0.0076

Correlation (inter-dependence) between parameters
Matrix of paired correlation coefficient of parameters, r_{ij}

	β_1	β_2	β_3
β_1	1.0000E + 00	−9.9681E − 01	9.8629E − 01
β_2	−9.9681E − 01	1.0000E + 00	−9.9523E − 01
β_1	9.8629E − 01	−9.9523E − 01	1.0000E + 01

Quality of achieved curve-fitting
Statistical analysis and goodness-of-fit test of classical residuals

	Independ. variable	Response measured	Prediction calculated	Standard deviation	Bias	Classical residual
i	x	y	$\hat{y}$	$s(\hat{y})$	$h(\hat{y})$	$\hat{e}$
1	1	1.6700E + 01	1.6695E + 01	1.9847E − 0	−1.2022E − 04	5.0684E − 03
2	5	1.6800E + 01	1.6790E + 01	1.5842E − 0	−1.7692E − 05	1.0093E − 02
3	10	1.6900E + 01	1.6921E + 01	1.2380E − 0	6.4131E − 05	−2.1134E − 02
4	15	1.7100E + 01	1.7068E + 01	1.1210E − 0	9.6718E − 05	3.2219E − 02
5	20	1.7200E + 01	1.7232E + 01	1.1897E − 0	8.5504E − 05	−3.1663E − 02
6	25	1.7400E + 01	1.7415E + 01	1.3105E − 0	4.0018E − 05	−1.4804E − 02
7	30	1.7600E + 01	1.7619E + 01	1.3837E − 0	−2.4973E − 05	−1.9465E − 02
8	35	1.7900E + 01	1.7848E + 01	1.3861E − 0	−8.8089E − 05	5.1821E − 02
9	40	1.8100E + 01	1.8104E + 01	1.4192E − 0	−1.1967E − 04	−3.7685E − 03
10	50	1.8700E + 01	1.8709E + 01	2.7266E − 0	8.4264E − 05	−8.5853E − 03

Statistical analysis

Residual sum of squares, $U(\hat{b})$:	5.9861E − 03
Regression rabat, $100D^2$, [%]:	9.9838E + 01
Akaike information criterion, AIC:	−6.4642E + 01
Estimate of standard deviation of prediction, $s(\hat{y}/x)$:	2.9243E − 02

Goodness-of-fit test

Estimate of residual variance, $s^2(\hat{e})$:	8.5516E − 04
Estimate of residual standard deviation, $s(\hat{e})$:	2.9243E − 02

Quality of experimental data
Indication of influential points (outliers and leverages):

Point	Jackknife residual	Cook distance	Diagonal elements	Normalized distance	Likelihood distance
i	$\hat{e}_J$	D	H_{ii}	FDA	LDA
1	7.0904E − 01	1.5852E − 02	4.6060E − 01	1.3469E − 03	8.8043E − 03
2	7.9083E − 01	2.3346E − 02	2.9349E − 01	5.8963E − 03	9.4152E − 03
3	−7.8569E − 01	4.6314E − 02	1.7922E − 01	2.0878E − 02	1.4915E − 02
4	1.1740E + 00	8.1704E − 02	1.4695E − 01	4.4161E − 02	4.7807E − 02
5	−9.9223E − 01	9.2870E − 02	1.6550E − 01	3.9056E − 02	4.2378E − 02
6	−6.3466E − 01	2.6862E − 02	2.0083E − 01	9.4246E − 03	9.4039E − 03
7	−7.5215E − 01	5.4890E − 02	2.2387E − 01	1.5231E − 02	1.1329E − 02
8	1.4315E + 00	3.9119E − 01	2.2466E − 01	2.6786E − 01	9.9158E − 01
9	−2.4752E − 01	2.2310E − 03	2.3554E − 01	7.5113E − 04	8.9988E − 03
10	−5.3222E − 01	1.4630E + 00	8.6934E − 01	4.1484E − 04	6.9048E − 03

Map of parameter sensitivity in model

Parameter	Relative change	Total sensitivity	Relative change
j	$CjR(-5\%)$ [%]	Cj	$CjR(+5\%)$ [%]
1	−1.1930E − 08	1.0000E + 00	2.8004E − 08
2	−7.0302E + 00	3.4946E + 00	7.6700E + 00
3	−1.7824E + 01	4.5182E + 03	2.1104E + 01

equations (26) and (27), the standard deviation of prediction $s(y/x)$, the total bias of parameter estimates M, equation (11), a graph of the confidence interval of prediction.

The goodness-of-fit contains: the table of the prediction calculated $\hat{y}$, the standard deviation of prediction $s(\hat{y})$, the residual bias $h(\hat{y})$, equation (9), and classical residuals $\hat{e}$, equation (7). Statistical characteristics describe classical residuals: the residual bias $E(\hat{e})$, equation (8), the norm of residual bias $\| E \|$, the mean of absolute residuals $E(|\hat{e}|)$, the mean of absolute values of relative residuals $100E(|\hat{e}_{rel}|)$ in percents, the estimate of residual variance $s^2(\hat{e})$, equation (24) and its square-root the residual standard deviation $s(\hat{e})$.

Prediction ability of model proposed. Prediction ability of model can be classified by the following procedure: data are divided on two groups, M_1 with indices $i = 1, \ldots, n/2$ and M_2 with indices $i = n/2 + 1, \ldots, n$. Denote estimates of parameters made from points of subgroup M_1 as $\hat{b}(M_1)$ and from subgroup M_2 as $\hat{b}(M_2)$. Prediction ability of the model is expressed by criterion

$$K = \frac{U(\hat{b})}{\sum_{i \in M_1} [y_i - f(x_i; b(M_2)]^2 + \sum_{i \in M_2} [y_i - f(x_i; b(M_1)]^2} \quad (33)$$

The prediction ability of the model is higher when the criterion K is close to one. The mean quadratic error of prediction MEP, equation (28), can also be calculated. The lower the value of MEP the better is the prediction ability of the proposed model.

Quality of experimental data. For examination of quality experimental data an identification of influential points by regression diagnostics is applied: the Jackknife residuals $\hat{e}_J$, the Cook distance D, the diagonal elements of projection (hat) matrix H_{ii}, the test criterion DSF, the normalized distance FDA, and the likelihood distance LD, equations (31) and (32).

Map of parameter sensitivity in model. The total sensitivity C_{Cj} for all parameters β_j and the relative changes caused by 5% change of parameters β_j are computed. Characteristics C_{Cj} and their interpretation are described in a forthcoming book.[4]

Graph of regression curve. A graph of regression curve fitted through given experimental points with the 95% confidence bands and two plots of classical residuals give a graphical overview of fitness achieved: the plot of type II and the plot of type III.

Physical meaning of parameter estimates. In proposed models some restrictions of physical meaning are given on parameter estimates. For example, concentrations or molar absorption coefficients are defined in a range of positive numbers only.

Software

Program MINOPT from CHEMSTAT package carries out the numerical and statistical analysis of a non-linear regression model $f(x; \beta)$ with use of modified "double dog–leg" strategy. This program contains all the above mentioned criteria of nonlinear model quality.

Program MINOPT is a part of CHEMSTAT package and is available from authors on request.

Illustrative examples

For illustration of MINOPT statistical characteristics the example of Model 1 from paper[3] was recomputed. Selected outputs are shown in Table 2.

CONCLUSION

Many problems in the chemical laboratory can be reduced to the problem of finding a correct mathematical model and its unknown parameters. It may be carried out by minimizing a difference between experimental and calculated data. The variety of regression diagnostics introduced here serves as an efficient tool in search of true model.

REFERENCES

1. M. Meloun and J. Čermák, *Talanta*, 1984, **11**, 947.
2. M. Meloun, J. Havel and E. Högfeldt, *Computation of Solution Equilibria*, Ellis Horwood, Chichester, 1988.
3. J. Militký and M. Meloun, *Talanta*, 1992, **39**,
4. M. Meloun, J. Militký and M. Forina, *Chemometrics for Analytical Chemistry*, Vol. 1: Interactive Data Analysis on IBM PC and Vol. 2: Interactive Model Building on IBM PC, Ellis Horwood, Chichester, 1992.
5. A. R. Gallant, *Nonlinear Statistical Models*, Wiley, New York, 1987.
6. J. R. Donaldson and R. B. Schnabel, *Technometrics*, 1987, **29**, 67.
7. R. D. Cook *et al.*, *Biometrika*, 1986, **73**, 615.
8. M. J. Box, *J. Roy. Stat. Soc.*, 1971, **B32**, 171.
9. G. P. Clarke, *J. Amer. Statist. Assoc.*, 1987, **82**, 221.
10. J. Militký, J. Čáp and K. Květon, *Proc. 2nd Int. Conference on Statistics*, Tampere, 1987.