ISSN 1995-4239, Numerical Analysis and Applications, 2015, Vol. 8, No. 4, pp. 344–350. © Pleiades Publishing, Ltd., 2015.
Original Russian Text © G.I. Rudoy, 2015, published in Sibirskii Zhurnal Vychislitel'noi Matematiki, 2015, Vol. 18, No. 4, pp. 425–434.

Applying Monte Carlo Methods to Analysis of Nonlinear Regression Models

G. I. Rudoy*

Dorodnitsyn Computing Center, Russian Academy of Sciences, ul. Vavilova 40, Moscow, 119333 Russia

Received June 4, 2014; in final form, March 24, 2015

Abstract—This paper presents a criterion, called coefficient stability, for determining errors in the coefficients of nonlinear regression models describing inexact data. A method for the estimation of coefficient stability is also described. The criterion is illustrated by a computational experiment with data obtained by measuring the refractive index as a function of the wavelength in the 400–1000 nm band for a transparent polymer. The convergence of the criterion to a known analytical solution for the case of linear regression is also studied.

DOI: 10.1134/S1995423915040072

Keywords: *symbolic regression, nonlinear models, solution stability, transparent medium dispersion, Monte Carlo methods.*

1. INTRODUCTION

Symbolic regression is often used to construct expert-interpretable models [1–5]. As applied to experiments in the natural sciences, it is an attempt to reconstruct a functional relation between parameters obtained from measurements and those specified with a certain accuracy, e.g., thermionic emission current of an electric lamp versus cathode temperature $I_k(T)$ under constant geometry and potential difference of the system; cw laser radiation power versus output mirror reflection coefficient $W_l(R)$ under constant radiation mode structure and active medium's excitation, the refraction index of a material versus wavelength $n(\lambda)$ under constant temperature, etc. Consider the last-mentioned case in more detail.

The following should be taken into account in regression analysis of such experiments:

1. All parameters measured (and controlled) at each experimental point are determined with some accuracy (which is usually known); the absolute error σ_i of the corresponding parameter may vary considerably in the range under study. For instance, if a diffraction lattice is used as the spectral device for a specific wavelength λ_i when measuring $n_i(\lambda_i)$, we have $\frac{\sigma_i}{\lambda_i} \approx \text{const}$, and it is not correct to assume that the error in determining the wavelength is constant for measurements in a wide spectral range.

2. As a rule, the experiment is performed to measure a function of one variable, x, that is, a functional relation of the form $y(x, \boldsymbol{\omega})$, where $\boldsymbol{\omega}$ is a set of parameters that are kept constant with finite accuracy. In some cases, this fact should be taken into account when constructing a model. However, usually the effect of variations in the experimental conditions can be estimated by the expert beforehand to provide stability of the measurements. Otherwise, the characteristic being measured is a function of several variables. The results of the present paper can be generalized to this case.

*E-mail: 0xd34df00d@gmail.com

3. In most cases, the expert knows the form of the sought-for functional relation or can consider a few possible variants, which simplifies the problem of regression. Most important for the expert are not only optimal values of the coefficients in the regression formula, but also the variance of these coefficients, as well as the interrelation between the variance and the accuracy in determining the quantities that are measured (controlled) in the experiment. This is essential, especially in those cases where the regression model coefficients are directly related to the fundamental characteristics of the process under study and used to calculate, for instance, the efficient mass of electrons in a semiconductor, Debye's temperature, resonant frequency, optical transition decay, etc. Here the accuracy of measuring the corresponding material constants is determined by the accuracy of calculating the regression model coefficients.

To our knowledge, the problem of nonlinear regression has not yet been considered in a statement when not only optimal coefficients of the regression model, but their errors are also determined. Well-known are some theoretical results for the case of linear regression

$$y = ax + b;$$

when the variance of all experimentally measured values y_i of the dependent variable y is the same, $D(y_i) = \sigma^2$, and the values of the independent variable, x_i, are exactly known, $D(x) = 0$. Then, in the representation

$$y_i = a(x_i - \overline{x}) + b + \xi_i \mid i \in \{1, \ldots, \ell\},$$

where $\overline{x} = \frac{\sum_{i=1}^{\ell} x_i}{\ell}$, and $\xi_i \sim \mathcal{N}(0, \sigma^2)$, according to [6] the random quantities a and b are independent and normally distributed. In addition, their variances are as follows:

$$D(a) = \frac{\sigma^2}{\sum\limits_{i=1}^{\ell}(x_i - \overline{x})^2}, \tag{1}$$

$$D(b) = \frac{\sigma^2}{\ell}. \tag{2}$$

In this paper, a general method for determining the error in the coefficients of nonlinear regression is proposed. The error in the regression parameters versus the accuracy in determining the wavelength and refraction index is determined using the function $n(\lambda)$ for a transparent polymer as an example. Here we restrict our consideration to one independent variable, λ. The method can easily be extended to the case of several variables.

2. MAJOR HYPOTHESIS

Consider a set of ℓ experimental points. In the vicinity of each of these points, a probability of deviation of the experimental values of the variables to be measured from this point (for instance, due to the presence of random measurement errors) is specified. In other words, let there be a sample $D = \{(x_i, y_i)\}$, $i = 1, \ldots, \ell$, and for each pair (x_i, y_i), a known distribution of probabilities of deviations, $\Delta_i x$ and $\Delta_i y$, of the independent and dependent variables, $P(\Delta_i x) = P_i^x(x - x_i)$ and $P(\Delta_i y) = P_i^y(y - y_i)$, from the values x_i and y_i, respectively. The mean of the deviation is assumed to be zero and, hence, the experimental values x_i and y_i are assumed to be average. The probabilities $P_i^x(x - x_i)$ and $P_i^y(y - y_i)$ are typically assumed to be Gaussian, and the variance values σ_i^x and σ_i^y for them are considered to be known.

Consider a parametric regression model $y(x, \boldsymbol{\omega})$ (generated inductively [5] or proposed by an expert) whose vector of parameters $\boldsymbol{\omega}$ minimizes some functional S, for instance, the following root-mean-square deviation:

$$S = \sum_{i=1}^{\ell} \big(y(x_i, \boldsymbol{\omega}) - y_i\big)^2 \underset{\boldsymbol{\omega}}{\to} \min. \tag{3}$$

For such a functional and its modifications taking into account the regression model's complexity [5], the procedure of minimization can be efficiently performed with the Levenberg–Marquardt algorithm (LMA) [7, 8].

Next we fix the structural form of the above function $y(x, \boldsymbol{\omega})$, and repeat the following computation procedure several times:

1. At the kth step, generate a random sample $D_k = \{(x_i^k, y_i^k)\}$, $i = 1, \ldots, \ell$: Obtain the value x_i^k from the corresponding value x_i of the initial sample D by adding random noise distributed according to P_i^x:

$$x_i^k = x_i + \xi_i^k, \quad \xi_i^k \sim P_i^x.$$

   Similarly, obtain y_i^k.

2. For the thus constructed data set or realization D_k, use the same optimization algorithm to find an optimal (minimizing the above functional) set $\boldsymbol{\omega}^k$ of regression coefficients $y(x, \boldsymbol{\omega})$ for the kth realization.

   Thus, for each regression coefficient $\boldsymbol{\omega}_\mathrm{p}$ we obtain a set of its values in the thus generated realizations $\{\boldsymbol{\omega}_\mathrm{p}^k\}$.

3. For a sufficiently large number of realizations M, the mean and standard deviation of the corresponding regression coefficient $\boldsymbol{\omega}_\mathrm{p}$ are defined as follows:

$$\overline{\boldsymbol{\omega}_\mathrm{p}} = \frac{\sum\limits_{i=1}^{M} \boldsymbol{\omega}_\mathrm{p}^i}{M}, \tag{4}$$

$$D(\boldsymbol{\omega}_\mathrm{p}) = \sigma_{\boldsymbol{\omega}_\mathrm{p}}^2 = \frac{1}{M-1} \sum_{i=1}^{M} (\boldsymbol{\omega}_\mathrm{p}^i - \overline{\boldsymbol{\omega}_\mathrm{p}})^2. \tag{5}$$

Our hypothesis is that the values obtained according to (4) and (5) are real. It is evident that the above-proposed approach to determining the error of regression coefficients is actually an application of a method of the Monte Carlo type to the regression problem.

From the above interpretation there also follows an evident criterion of the computation procedure termination when with increasing number of realizations, M, the variation in the values $\overline{\boldsymbol{\omega}_\mathrm{p}}$ and $D(\boldsymbol{\omega}_\mathrm{p})$ becomes less than a value chosen by the expert.

It should be noted that in the general case the limits of expressions of the type (4) and (5) as $M \to \infty$ may not exist, which makes the above-proposed computational scheme incorrect. For sufficiently smooth functions that are of practical interest to us, correctness of the procedure can be rigorously proved, which, however, is beyond the scope of our paper.

3. MODEL EXAMPLE

Let us test the method using a known analytical solution of (1), (2) by calculating the errors of the coefficients of a linear relation with an exactly known independent variable and a Gaussian distribution of the error in the dependent variable. In this case the parameters of error distribution in the dependent variable are the same for each experimental point. In the computational experiment, the relation being simulated has the form $y = ax + b$, $a = 3$, $b = 10$. The independent variable is defined at $\ell = 10$ (in another experiment $\ell = 50$) points of the interval $[0, 10]$, and the number of realizations is 100 million. The regression coefficients are optimized using the Levenberg–Marquardt algorithm, by analogy with a more general case of nonlinear regression.

In Figs. 1–3, the number of realizations N is shown on the x axis, and the relation $\delta = \frac{\sigma_\mathrm{ce} - \sigma_\mathrm{ex}}{\sigma_\mathrm{ex}}$ for the coefficients a, b, on the y axis. Here σ_ce is the variance obtained in the computational experiment, and σ_ex is the exact theoretical variance value according to (1), (2).

One can see that already for $N \approx 2.5 \cdot 10^7$ the relative difference $\delta = \frac{\sigma_\mathrm{ce} - \sigma_\mathrm{ex}}{\sigma_\mathrm{ex}}$ does not exceed 0.05% regardless of the number of points at which the independent variable is defined and the variance of the random quantity. This seems to be a good result, which proves that the approach being discussed and the main hypothesis are correct.

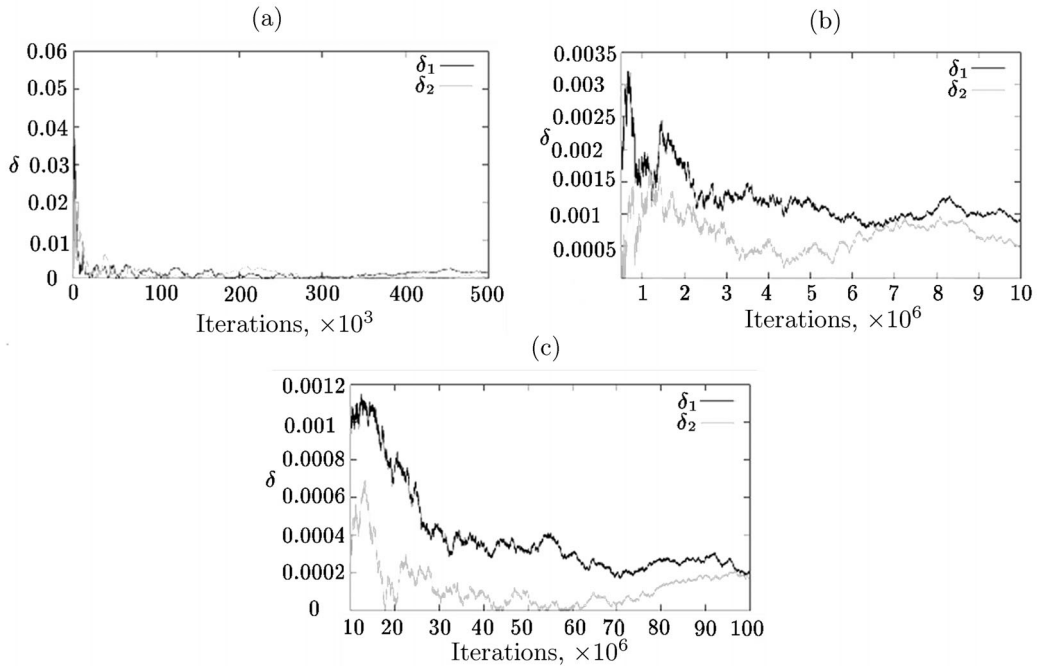

Fig. 1. δ versus the number of iterations N at $D(\xi) = 10$ and $\ell = 10$; (a) $N \in [0; 5 \cdot 10^5]$, (b) $N \in [5 \cdot 10^5; 10^7]$, (c) $N \in [10^7; 10^8]$.

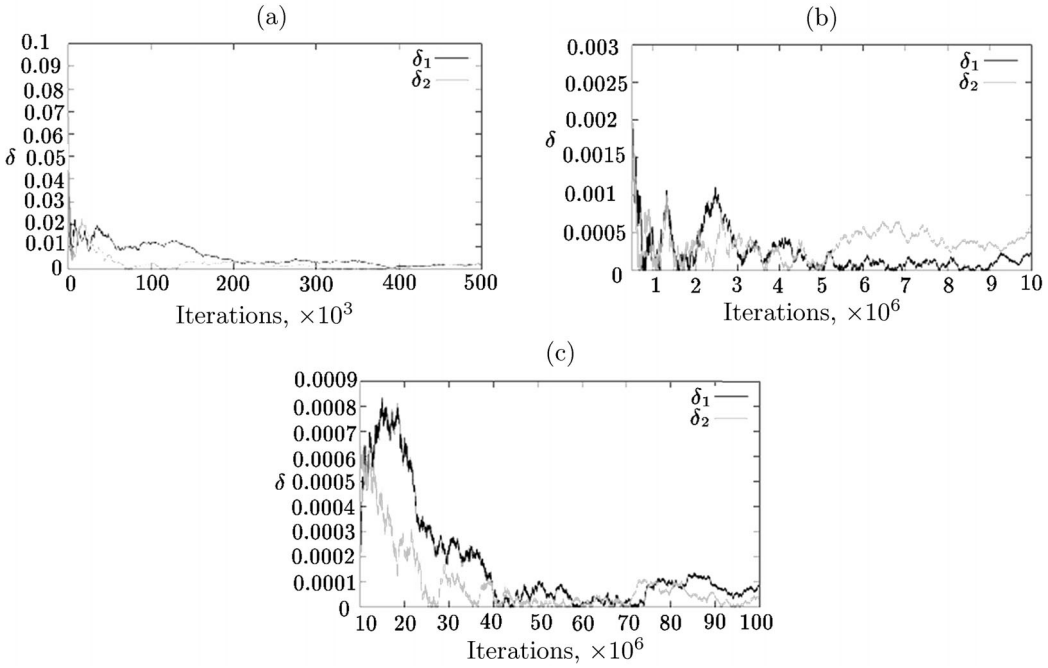

Fig. 2. δ versus the number of iterations N at $D(\xi) = 1$ and $\ell = 10$; (a) $N \in [0; 5 \cdot 10^5]$, (b) $N \in [5 \cdot 10^5; 10^7]$, (c) $N \in [10^7; 10^8]$.

4. ERROR IN THE REGRESSION COEFFICIENTS OF $n(\lambda)$

In the experiment, 17 values of the wavelength were used, and the refraction index of a transparent (in the spectral range under study) polymer was measured. The results are presented in Table 1.

The absolute error in measuring the refraction index σ_n is typically $(3 \div 10) \cdot 10^{-5}$ in the entire range of wavelengths (λ, nm), and the relative error in determining the wavelength σ_λ/λ for the diffraction devices in question is $(3 \div 30) \cdot 10^{-4}$.

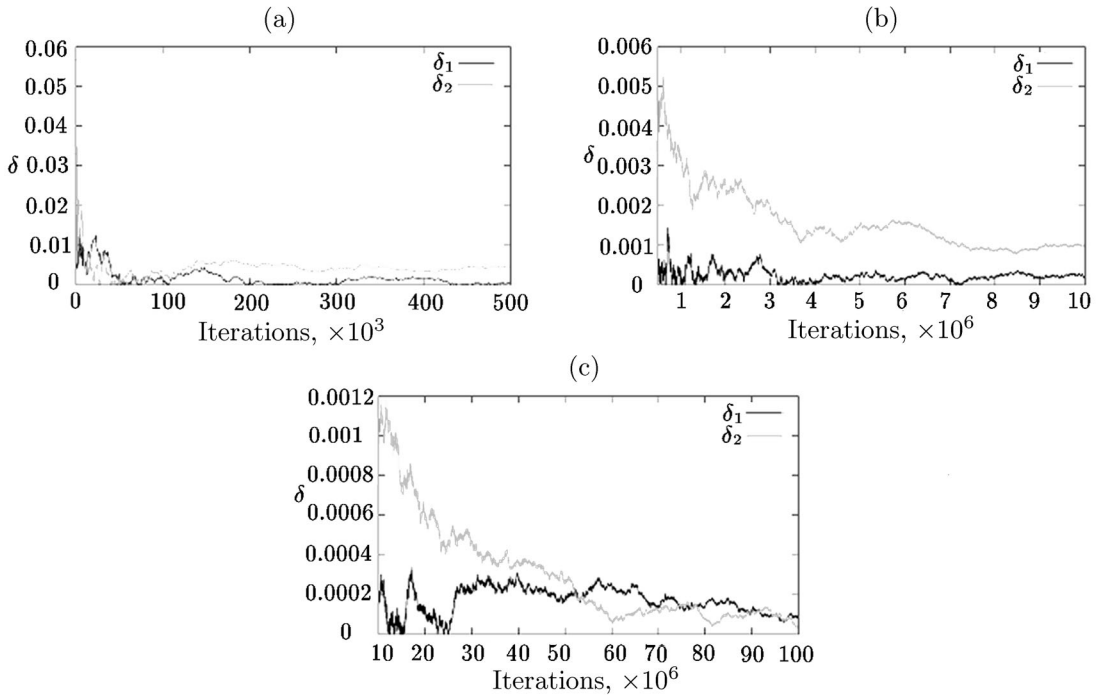

Fig. 3. δ versus the number of iterations N at $D(\xi) = 1$ and $\ell = 50$; (a) $N \in [0; 5 \cdot 10^5]$, (b) $N \in [5 \cdot 10^5; 10^7]$, (c) $N \in [10^7; 10^8]$.

Table 1. Experimental values of the refraction coefficients

λ, nm	$n(\lambda)$	λ, nm	$n(\lambda)$	λ, nm	$n(\lambda)$
435.8	1.35715	577.0	1.34968	750	1.34607
447.1	1.35625	587.6	1.34946	800	1.34544
471.3	1.35449	589.3	1.34938	850	1.34487
486.1	1.35349	656.3	1.34768	900	1.34437
501.6	1.35275	667.8	1.34740	950	1.34407
546.1	1.35083	706.5	1.34664		

The model of $n(\lambda)$ proposed by an expert is to determine the error of regression coefficients caused by the experimental errors in the measurement of the wavelength and refraction index.

In addition, it is reasonable to find the relationship between the error in the regression coefficients and the accuracy of the experiment, which is of both theoretical and practical interest. For instance, an increase in the measurement accuracy of the refraction index n of up to $2 \cdot 10^{-5}$ requires much effort and time. However, if the error in determining the regression model coefficients changes only slightly, further modernization of the experimental setup does not make sense. Similarly, the question of what increase in the accuracy of determining the wavelength is necessary to considerably decrease the error in determining the regression coefficients is of both theoretical and practical significance for increasing the accuracy in determining the refraction index.

A functional relation of the following form was proposed by an expert:

$$n(\lambda) = a + \frac{b}{\lambda^2} + \frac{c}{\lambda^4}. \tag{6}$$

For this type of nonlinear regression, the LMA was used to determine optimal regression coefficients

Table 2. Relative error in determining the regression coefficient a

σ_λ/λ \ σ_n/n	$2 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$3 \cdot 10^{-4}$
10^{-3}	$1.255 \cdot 10^{-5}$	$5.54 \cdot 10^{-5}$	$1.656 \cdot 10^{-4}$
10^{-2}	$6.32 \cdot 10^{-5}$	$8.24 \cdot 10^{-5}$	$1.753 \cdot 10^{-4}$

Table 3. Relative error in determining the regression coefficient b

σ_λ/λ \ σ_n/n	$2 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$3 \cdot 10^{-4}$
10^{-3}	$2.14 \cdot 10^{-3}$	$8.68 \cdot 10^{-3}$	$2.54 \cdot 10^{-2}$
10^{-2}	$1.36 \cdot 10^{-2}$	$1.59 \cdot 10^{-2}$	$2.85 \cdot 10^{-2}$

Table 4. Relative error in determining the regression coefficient c

σ_λ/λ \ σ_n/n	$2 \cdot 10^{-5}$	$1 \cdot 10^{-4}$	$3 \cdot 10^{-4}$
10^{-3}	1.15	1.426	1.454
10^{-2}	1.45	1.456	1.47

providing the root-mean-square error $\approx 3.97 \cdot 10^{-9}$ for $a = 1.33344$; $b = 2841.63$; $c = 1599.27$; the wavelength is measured in nanometers.

The errors in the regression model coefficients were calculated for some combinations of errors in determining the wavelengths (in relative units) and refraction index (also in relative units). In each case, statistical processing of a million of realizations was performed. Tables 2–4 present some results of the computational experiment.

For $\frac{\sigma_\lambda}{\lambda} < 10^{-4}$, the errors in the first and second regression coefficients depend but slightly on further increase in the accuracy of determining the wavelength, and almost linearly depend on the accuracy of measuring the refraction index. The absolute error of determining the third regression coefficient is great and comparable to the coefficient itself. This is, evidently, due to the fact that the contribution of the third term in (6) to the refraction index is minor: changing this term only slightly affects the root-mean-square deviation. Hence, this component is determined with a low accuracy.

5. CONCLUSIONS

1. The method proposed in this paper to determine errors in the regression coefficients can be used at any probability distribution of the error in the dependent and independent variables, including error distributions for various variables and for a single variable at various experimental points.

2. The method will give somewhat different error values when different functionals of the error, S, are used. This situation is not unusual [5]. For instance, the regression coefficients will be different if the sum of squared distances along an axis corresponding to the dependent variable from the experimental points to the approximating regression model (according to (3)) is minimized by the standard least squares method, or if the sum of Euclidean distances from the experimental points to the approximating regression model is minimized (as in the well-known paper by K. Pearson [9]).

In our opinion, the optimal functional G should be chosen by the expert, first of all, by analyzing the errors in determining the dependent and independent variables. Specifically, the least-squares method is preferable if the independent variable is determined exactly, and the Pearson approach is preferable if the errors in the variables are close to each other.

3. To optimize the regression coefficients, the LMA was used in standard form when a functional of the type (3) is minimized. As shown above, this approach is correct when the error in determining the independent variable is small. In the general case, it is more correct to minimize the sums of squares of the regression residuals corresponding to the distance to the regression curve (see [9]) taking into account the errors in the variables being measured, but not the distance to the point in the curve with a fixed coordinate x, which is correct when there is no error in the measurement of the independent variable. For sufficiently smooth regression relations, when the relation in the vicinity of (x_i, y_i) can be linearized, the corresponding distance has the form $\frac{(y_i - f(x_i, \boldsymbol{\omega}))^2}{(\sigma_i^y)^2 + k^2(\sigma_i^x)^2}$, where $k = \frac{\partial f}{\partial x}|_{x=x_i}$. In other words, the "standard" regression residual is normalized by some quantity depending on the measurement errors σ_i^x and σ_i^y and the regression curve slope in the vicinity of the point (x_i, y_i) in question. A corresponding modification of the LMA will be the subject of a future study.

For the computational experiments whose results are presented in the present paper, the use of the standard LMA can be justified. Specifically, for the model case considered in Section 3 it is assumed in advance that the independent variable is known exactly, $\forall i : \sigma_i^x = 0$, and the errors in determining the dependent variable are constant: $\sigma_i^y = \text{const}$. Then the denominator of the above expression for the distance reduces to a constant by which every regression residual is normalized. This complies with the condition of applicability of the exact result (1), (2).

In Section 4, the regression relation for the dispersion of a polymer was analyzed for the errors σ_n and σ_λ (σ^y and σ^x, respectively) in the entire spectral range at $k = \frac{\partial n}{\partial \lambda} \leq \frac{1}{5}\frac{\sigma_n}{\sigma_\lambda}$. This allows using the standard LMA with a sufficient accuracy.

Thus, in this paper a method for determining errors in the regression model coefficients, including the case of nonlinear regression, was proposed. In a computational experiment, a good agreement with a theoretical result when the standard deviation of the regression coefficients is determined exactly has been obtained. It has been shown that the computational algorithm can be used in the analysis of specific physical experiments.

REFERENCES

1. Davidson, J.W., Savic, D.A., and Walters, G.A., Symbolic and Numerical Regression: Experiments and Applications, *Develop. Soft Computing*, 2001, vol. 9, pp. 175–182.
2. Sammut, C. and Webb, G.I., *Symbolic Regression, Encyclopedia of Machine Learning*, Berlin: Springer-Verlag, 2010; http://dx.doi.org/10.1007/978-0-387-30164-8.
3. Strijov, V. and Weber, G.W., Nonlinear Regression Model Generation Using Hyperparameter Optimization, *Math. Appl.*, 2010, vol. 60, no. 4, pp. 981–988, http://dx.doi.org/10.1016/j.camwa.2010.03.021.
4. Strizhov, V.V., Methods of Inductive Generation of Regression Models, Preprint of the Dorodnitsyn Computing Center, Russ. Acad. Sci., Moscow, 2008.
5. Rudoy, G.I. and Strizhov, V.V., Algorithms for Inductive Generation of Superpositions for Approximation of Experimental Data, *Inform. Appl.*, 2013, vol. 7, no. 1, pp. 44–53.
6. Vatutin, V.A., Ivchenko, G.I., Medvedev, Yu.I., and Chistyakov, V.P., *Teoriya veroyatnostei i matematicheskaya statistika v zadachakh* (Probability Theory and Mathematical Statistics in Exercises), 3rd ed., Moscow: Drofa, 2005.
7. Marquardt, D.W., An Algorithm for Least-Squares Estimation of Non-Linear Parameters, *J. Soc. Industr. Appl. Math.*, 1963, vol. 11, no. 2, pp. 431–441.
8. More, J.J., The Levenberg–Marquardt Algorithm: Implementation and Theory, *Numerical Analysis, Lect. Notes Math.* 1978, vol. 630, pp. 105–116.
9. Pearson, K., On Lines and Planes of Closest Fit to Systems of Points in Space, *Philos. Mag.*, 1901, vol. 2, no. 6, pp. 559–572.