ITERATED NON-LINEAR REGRESSION

*Miroslav Svítek**

Abstract: The paper presents iterated algorithm for parameter estimation of non-linear regression model. The non-linear model is firstly approximated by a polynomial. Afterwards, parameter estimation based on measured data is taken as the initial value for the proposed iterated algorithm. As the estimation method, the well-known Least Square Estimation (LSE), artificial neural networks (ANN) or Bayesian methodology (BM) can be used. With respect to the knowledge of initial parameters the measured data are transformed to meet best the non-linear regression criteria (orthogonal data projection). The original and transformed data are used in the next step of the designed iterated algorithm to receive better parameter estimation. The iteration is repeated until the algorithm converges into a final result. The proposed methodology can be applied on all non-linear models that could be approximated by a polynomial function. The illustrative examples show the convergence of the designed iterated algorithm.

Key words: *parameter estimation, adaptive systems, dynamical modeling, iterated estimation, polynomial regression, Least Square Estimation, Bayesian methodology, artificial neural networks*

Received: January 13, 2014 **DOI:** 10.14311/NNW.2014.24.024
Revised and accepted: August 20, 2014

1. Introduction

There is a number of problems available to estimate unknown parameters of non-linear models as common statistical methods are primarily designed for linear equations. References [8, 7] speak about a non-correct task referring to parameter estimation of a non-linear regression model. Publications [1, 5] provide a comprehensive reference on non-linear regression and non-linear least squares estimation where unknown coefficients may be estimated from a linearized version of a model. Some other approaches are based on application of kernel density non-parametric estimators [9, 12] with guaranteed statistical features [6].

The main idea of the presented approach is not to change the regression model itself (e.g. its linearization) but to transform the measured data through orthogonal projection. The orthogonal data projection onto non-linear regression function is strongly dependent on unknown parameters. On the other hand, unknown parameters must be estimated from the measured data. This logical circle brings complexity into the designed non-linear estimation algorithm.

*Miroslav Svítek
Faculty of Transportation Sciences, Czech Technical University in Prague, Czech Republic, E-mail: svitek@fd.cvut.cz

Nevertheless, if the non-linearity of the regression model was approximated by a polynomial function, the designed iterated algorithm could be newly introduced. The iterated approach yields into modification of a well-known estimation algorithm, such as Least Square Estimation (LSE), artificial neural networks (ANN), Bayesian methodology (BM), etc. The presented iteration of these algorithms offers a better parameter estimation of a polynomial regression model.

The paper is structured as follows. Paragraph 2 defines mathematical background of the presented approach dealing with the orthogonal data projection and its application to a polynomial regression model. Paragraph 3 introduces the designed iterated algorithm. Paragraph 4 concerns itself with the theoretical results and Paragraph 5 explains the application of the designed algorithm by way of two illustrative examples. Paragraph 6 then concludes the paper.

2. Mathematical background

2.1 Orthogonal data projection

Let us define the non-linear function $f(a, x)$ that depends on unknown parameters a. The principle of iterated parameter estimation comes from the orthogonal data projection to the non-linear function as it is shown in Fig. 1.

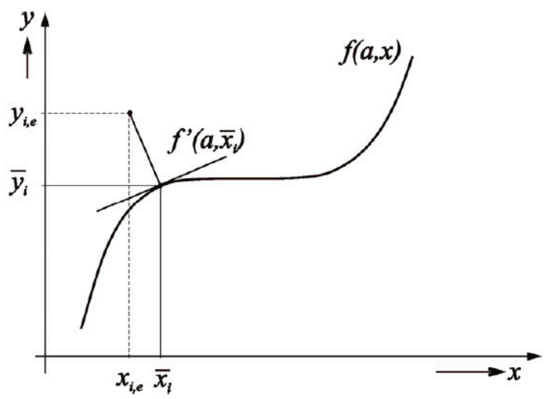

Fig. 1 *Orthogonal data projection in non-linear function.*

Let us suppose the availability of the noisy experimental data $\{x_{i,e}, y_{i,e}\}$ where $i \in \{1, 2, \ldots, N\}$. It is reasonable to anticipate the noise distortion in both x and y axes. For ideal situation (without an added noise signal) we may expect data $\{\bar{x}_i, \bar{y}_i\}$ that can be taken as mean values of the experimental data.

The orthogonal projection of the experimental data $\{x_{i,e}, y_{i,e}\}$ to a non-linear function $f(a, x)$ depends on the derivation $f'(a, \bar{x}_i)$ in the point $\bar{x}_i$. Unfortunately, we know neither the unknown parameters a nor the point $\bar{x}_i$. With respect to geometry shown in Fig. 1 we can note down the following equation:

$$f'(a, \bar{x}_i) = \frac{x_{i,e} - \bar{x}_i}{y_{i,e} - f(a, \bar{x}_i)}. \tag{1}$$

A new function related to (1) can be introduced:

$$\chi_i\left(\bar{x}_i\right) = \bar{x}_i - x_{i,e} + f'\left(a, \bar{x}_i\right) \cdot \left[y_{i,e} - f\left(a, \bar{x}_i\right)\right]. \tag{2}$$

This function must fulfill the condition

$$\chi_i\left(\bar{x}_i\right) \equiv 0 \tag{3}$$

to guarantee the best orthogonal projection.

2.2 Orthogonal projection for polynomial functions

In this paragraph we suppose that a non-linear function $f\left(a, x\right)$ is approximated in the given interval by the polynomial

$$f\left(a, x\right) \approx a_0 + a_1 \cdot x + a_2 \cdot x^2 + \cdots + a_M \cdot x^M \tag{4}$$

where the non-linear function approximation is out of the paper's scope, nevertheless, a lot of references are available, e.g. [11] which presents how to estimate the integer parameter M in (4).

The equation (2) can be easily rewritten for polynomial functions as follows:

$$\begin{aligned} \chi_i\left(\bar{x}_i\right) &= \bar{x}_i - x_{i,e} + + \left(a_1 + 2 \cdot a_2 \cdot \bar{x}_i + 3 \cdot a_3 \cdot \bar{x}_i^2 + \cdots + M \cdot a_M \cdot \bar{x}_i^{M-1}\right) \cdot \\ &\quad \cdot \left(y_{i,e} - a_0 - a_1 \cdot \bar{x}_i - a_2 \cdot \bar{x}_i^2 - \cdots - a_M \cdot \bar{x}_i^M\right) \end{aligned} \tag{5}$$

It is evident that the equation (5) is *2M-1* polynomial of an unknown variable $\bar{x}_i$. The coefficients of this polynomial could be computed by using experimental data $\{x_{i,e}, y_{i,e}\}$ and the parameters $\{a_0, a_1, \ldots, a_M\}$. We can now assume that we have an up-to-date parameter estimate $\{\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_M\}$ at our disposal which could be estimated by means of the measured data $\{x_{i,e}, y_{i,e}\}$. A number of algorithms for parameter estimation is available, e.g. Least Square Estimation (LSE), Bayesian methodology, etc. [10]. However, these algorithms do not expect noise distortion in both x and y axes and due to non-linearity the accuracy of parameter estimation is not accurate enough. On the other hand, such imprecise estimation can be used as an initial value for the presented iterated method.

The solution of the equation (3) represents the roots of the *2M-1* polynomial (5). This task is feasible because there is a number of numerical algorithms to tackle this problem. Numerical methods give us a set of different roots. It should be mentioned that equation (3) has at least one real root as it is a polynomial of the odd order.

The next step of this methodology is selecting the best root of the polynomial (5) that will represent the appropriate transformed data $\{\bar{x}_i, \bar{y}_i\}$ where $\bar{y}_i = f\left(\hat{a}, \bar{x}_i\right)$. Let us define the following two conditions for an appropriate root selection:

- the selected root must be real,
- the selected root must fulfil the condition of minimal length between transformed and measured y-values $|\bar{y}_i - y_{i,e}|$.

These two conditions guarantee that the transformed data $\{\bar{x}_i, \bar{y}_i\}$ will satisfy the subsequent processing of the designed iterated algorithm.

3. Iterated estimation algorithm

Referring to the mathematical background given in the previous paragraph we shall try to design the numerical algorithm in the following three steps:

A. Initial step Let there be N-samples of the measured data $\{x_{i,e}, y_{i,e}\}$.

The estimation algorithm (LSE, artificial neural networks, Bayesian approach, etc.) is applied to estimate the initial parameters $\{\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_M\}$ of the polynomial model:

$$y_{i,e} \approx \hat{a}_0 + \hat{a}_1 \cdot x_{i,e} + \hat{a}_2 \cdot x_{i,e}^2 + \cdots + \hat{a}_M \cdot x_{i,e}^M. \tag{6}$$

B. Data transformation We take actual the estimated parameters $\{\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_M\}$ together with the available experimental data $\{x_{i,e}, y_{i,e}\}$ and use them in an numerical algorithm for finding the roots of a polynomial (different possible variants of unknown data $\bar{x}_i$):

$$\begin{aligned} \bar{x}_i - x_{i,e} + \left(\hat{a}_1 + 2 \cdot \hat{a}_2 \cdot \bar{x}_i + 3 \cdot \hat{a}_3 \cdot \bar{x}_i^2 + \cdots + M \cdot \hat{a}_M \cdot \bar{x}_i^{M-1}\right) \cdot \\ \cdot \left(y_{i,e} - \hat{a}_0 - \hat{a}_1 \cdot \bar{x}_i - \hat{a}_2 \cdot \bar{x}_i^2 - \cdots - \hat{a}_M \cdot \bar{x}_i^M\right) \equiv 0. \end{aligned} \tag{7}$$

As there are real or complex roots of a polynomial (7) available we select the most appropriate one with respect to the two conditions introduced in paragraph 2.2.

This part of algorithm yields into the transformed data $\{\bar{x}_i, \bar{y}_i\}$ where $\bar{y}_i = f(\hat{a}, \bar{x}_i)$.

C. Iterated estimation of parameters We compute a new data set $\{x_{i,a}, y_{i,a}\}$ as an average of transformed and measured data,

$$x_{i,a} = \frac{x_{i,e} + \bar{x}_i}{2}, \tag{8}$$

$$y_{i,a} = \frac{y_{i,e} + \bar{y}_i}{2}. \tag{9}$$

The new iteration of the estimation algorithm (LSE, Bayesian approach, etc.) is applied on average data $\{x_{i,a}, y_{i,a}\}$ to estimate the new parameters $\{\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_M\}$ of the polynomial model:

$$y_{i,a} \approx \hat{a}_0 + \hat{a}_1 \cdot x_{i,a} + \hat{a}_2 \cdot x_{i,a}^2 + \cdots + \hat{a}_M \cdot x_{i,a}^M. \tag{10}$$

The newly estimated parameters $\{\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_M\}$ are used as a new input into the data transformation procedure (phase B).

Continual repetition of the actual data transformation (phase B) and the procedure of a new parameter estimation (phase C) results in iterated estimation algorithm of model parameters.

4. Discussion

If the noise signal is only added to the function $f(a, x)$ (no distortion in x-coordinate) the iterated algorithm will not aid well and iterations will theoretically show similar results. With the noise added to the x-coordinate the transformed noise signal strictly depends on x-values.

The initial estimate de facto supposes no added noise to the x-coordinate. The subsequent iteration tries to correct the non-linear dependence on x-values so the new estimation has either to improve, or in the worst case remain the same. Thus the algorithm converges into improved estimated values of unknown parameters.

It is evident that a step-by-step correction of non-linearity yields gradually into a final orthogonal data projection presented in Fig. 1. This means that corrections of the parameter estimates get lower and lower. As soon as the correction between the current and the last parameter estimate becomes lower than the predefined value (stopping criteria) the algorithm is stopped.

5. Illustrative examples

In this paragraph the designed algorithm is shown on two illustrative examples with predefined polynomial dimension (chosen parameter M). The first example demonstrates the convergence of the iterated algorithm for one-dimensional case, the second applies to a more complex non-linear regression model.

We will study the convergence of the iterated estimation process for an example where the non-linear model is not a polynomial. The theoretical analysis of parameter M estimation or practical recommendation obtained e.g. by numerical experiments will be presented.

Example 1: Let us define the quadratic non-linear model

$$y_i = a \cdot x_i^2 \tag{11}$$

with the parameter $a = 0.1$ and the following added noise-signals

$$y_i = \bar{y}_i + e_{y,i}, \tag{12}$$

$$x_i = \bar{x}_i + e_{x,i}. \tag{13}$$

We suppose independent noises with following parameters:

$$e_{y,i} \approx N\left(0,\ 10\right), \tag{14}$$

$$e_{x,i} \approx N\left(0,\ 5\right). \tag{15}$$

Taking into account all the above mentioned information we can rewrite the non-linear model as follows:

$$\bar{y}_i = a \cdot \left(\bar{x}_i + e_{x,i}\right)^2 + e_{y,i}. \tag{16}$$

First of all, the model (11) is used to generate 100 equidistant data $\{\bar{x}_i, \bar{y}_i\}$from the interval $\langle 0,\ 15\rangle$(shown as a continuous line in Fig. 2) and then the noise signals are added to simulate the noisy data $\{x_i, y_i\}$ (printed as '*' points in Fig. 2).

The convergence of the iterated parameter estimation algorithm is demonstrated in Fig. 3. Fig. 4 presents the error comparison (quality of estimation algorithm) between classical LSE (initial LSE estimation) and the newly developed iterated LSE algorithm. It is evident that the new iterated LSE algorithm

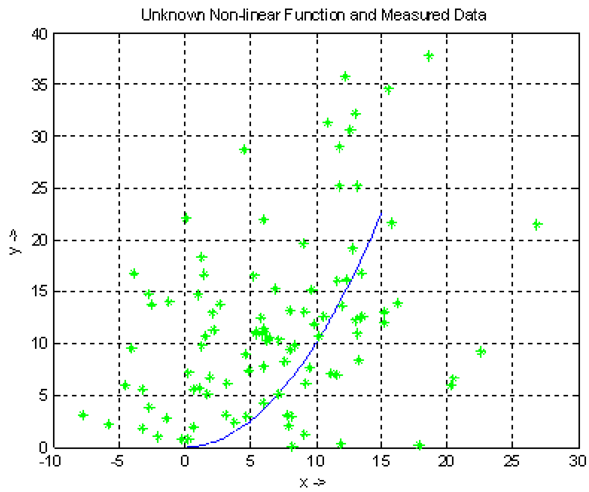

Fig. 2 *Generated Ideal (continuous line) and Noisy Data (printed as '*' points).*

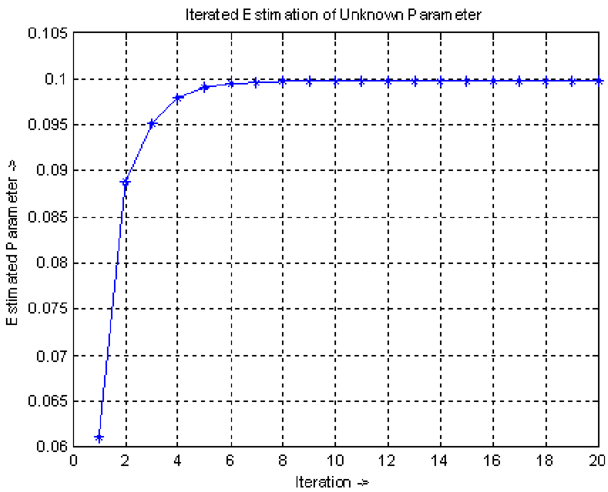

Fig. 3 *Iterated Algorithm of Parameter Estimation.*

can eliminate the error signal better, the error function (after 20 iterations) being more equally distributed – relative to the first (initial) estimate.

Considering a hundred of different realizations of Iterated LSE algorithm presented in Fig. 5 we can achieve the following performance parameters: the mean value of estimated parameter $\hat{a} = 0.0989$and a standard deviation $\sigma_{\hat{a}} = 0.0132$.

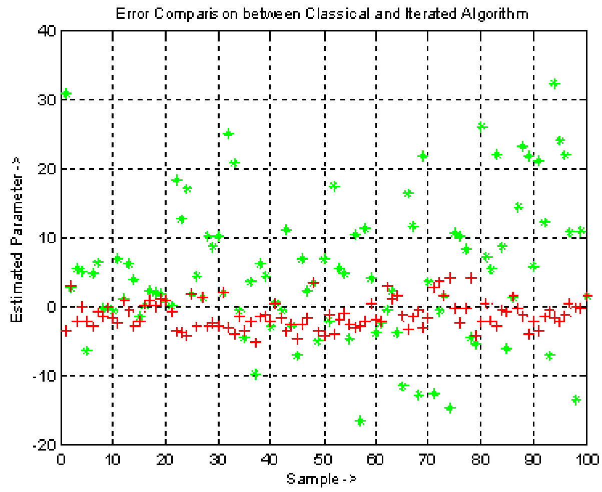

Fig. 4 *Error Comparison between Classical (error signal printed as '*') and Iterated LSE Algorithms (printed as '+').*

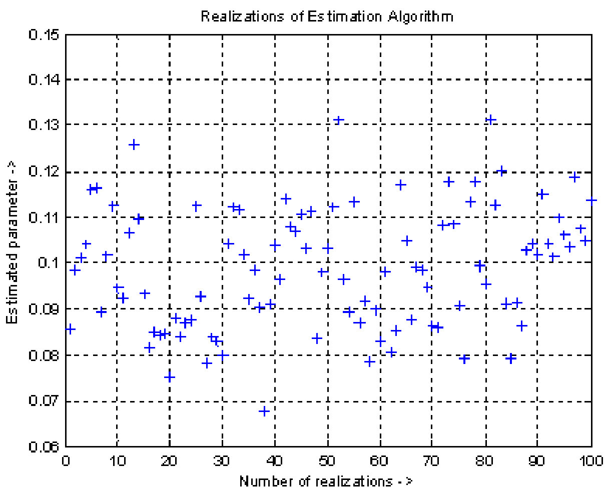

Fig. 5 *Different realizations of Iterated LSE algorithm for algorithm performance assessment.*

Example 2: Let us define a more complex non-linear model

$$y_i = a \cdot x_i^2 + b \cdot x_i^3 \tag{17}$$

with parameters $a = 5$, $b = -3$ and added noise-signals

$$e_{y,i} \approx N\left(0,\, 5\right), \tag{18}$$

$$e_{x,i} \approx N\left(0,\, 2\right). \tag{19}$$

We generate 100 equidistant data $\{\bar{x}_i, \bar{y}_i\}$ from the interval $\langle 0,\, 15\rangle$ (shown as a continuous line in Fig. 6) and then the noisy data $\{x_i, y_i\}$ (printed as '*' points in Fig. 6).

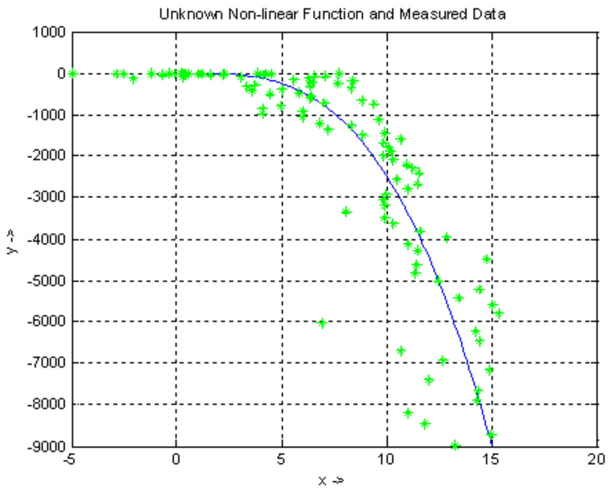

Fig. 6 *Generated Ideal (continuous line) and Noisy Data (printed as '*' points).*

The convergence of both estimated parameters is shown in Fig. 7. This example proves the numerical feasibility and practical applicability of the presented methodology.

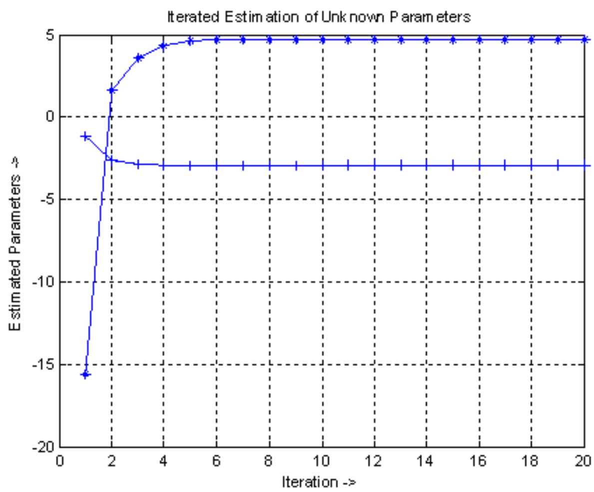

Fig. 7 *Iterated Algorithm of Parameter Estimation.*

If we perform 100 of different realizations of Iterated LSE algorithm under similar conditions we can compute the following statistical parameters: the mean value of estimated parameters $\hat{a} = 4.654$ $\hat{b} = -2.9632$ and standard deviations $\sigma_{\hat{a}} = 1.1768$ $\sigma_{\hat{b}} = 0.1040$.

6. Conclusion

The goal of the paper was to present the basic idea and a new approach to the estimation of a non-linear regression model. In past a lot of approximation algorithms were designed to tackle non-linear problems. Such algorithms tried to more or less linearize the non-linear regression function. Our approach is different. We do not attempt to linearize the non-linear model itself but to design the data transformation with respect to the non-linear model.

It was proven on illustrative examples that such approach is numerically feasible and the new transformed data match the non-linear model even if the initial parameters are not exact. Suitable combination of the original (measured) and the transformed data can converge into better precision of an unknown parameter estimation.

The presented method can easily be extended to other estimation procedures, such as Bayesian methodology, artificial neural networks, etc. The approach will be similar, i.e. these methods can be used for the first estimation of unknown parameters and then implemented in the proposed transformation of the measured data. Combination of row and transformed data can then be used as average data for the next estimation procedure.

The presented iterated method will be further developed in the future research. In mathematical area the algorithm's performance parameters should be derived. From the numerical point of view the algorithm for multi-dimensional polynomial models will be tested. We believe that a modified iterated algorithm can also be used for solving other non-linear regression problems as well as it can contribute to further development of the complex dynamical systems [14, 4, 3, 13, 2].

References

[1] BATES D.M., WATTS D.G. *Nonlinear Regression Analysis and Its Applications.* New York: John Wiley & Sons, Inc., 1988, doi: 10.1002/9780470316757.

[2] BĚLINOVÁ Z. Self-organizing Aspects in Transportation and Their Modelling. *Neural Network World.* 2008, 18(1), pp. 33–38.

[3] BOUCHNER P., HAJNÝ M., NOVOTNÝ M., PIEKNÍK R., SOJKA J. Car Simulation and Virtual Environments for Investigation of Driver Behavior. *Neural Network World.* 2005, 15(2), pp. 149–163.

[4] BOUCHNER P., NOVOTNÝ S., FABER J., TICHÝ T. Driver's Attention Level Improvement with Use of Biofeedback Stimulation Incorporated into Driving Simulator. *Neural Network World.* 2009, 19(1), pp. 109–118.

[5] GALLANT A.R. Nonlinear Regression. *The American Statistician.* 1975, 29(2), pp. 73–81.

[6] MACHKOURI M. Asymptotic normality of the Parzen–Rosenblatt density estimator for strongly mixing random fields. *Statistical Inference for Stochastic Processes.* 2011, 14(1), pp. 73–84, doi: 10.1007/s11203-011-9052-4.

[7] OSOVETS S.V., AZIZOVA T.V., DAY R.D., WALD N., DRUZHININA M.B. Assessment of risks and dose thresholds for some effects of acute exposure. *Health Physics.* 2011, 100(2), pp. 176–184, doi: 10.1097/HP.0b013e3181ebd123.

[8] OSOVETS S.V. Methods of orthogonal regression for some no correct tasks in biology. In: *Proceedings of the 4th International Conference on Mathematical Biology and Bioinformatics*, Pushchino, Russia (V.D. Lakhno Eds.). Moscow: MaxPress 2012, pp. 94–95.

[9] PARZEN E. On estimation of a probability density function and mode. *The annals of mathematical statistics.* 1962, 33, pp. 1065–1076.

[10] PETERKA V. *Bayesian Approach to System Identification.* In P. Eykhoff (Ed.), Trends and Progress in System Identification. New York: Pergamon, Elmsdorf, 1981, pp. 239–304, doi: 10.1016/B978-0-08-025683-2.50013-2.

[11] PHILLIPS G.M. *Interpolation and Approximation by Polynomials.* CMS Books in Mathematics. New York: Springer, 2003, doi: 10.1007/b97417.

[12] ROSENBLATT M. A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America.* 1956, 42(1), pp. 43–47, doi: 10.1073/pnas.42.1.43.

[13] SADIL J. Artificial intelligence methods used for predicting traction power consumption. *Neural Network World.* 2008, 18(1), pp. 15–22.

[14] SVÍTEK M. Applying Wave Probabilistic Functions for Dynamic System Modeling. *IEEE Transactions on Systems, Man and Cybernetics. Part C-Applications and Reviews.* 2011, 41(5), pp. 674–681, doi: 10.1109/TSMCC.2010.2093127.