UDC 519.2+004.8 https://doi.org/10.20535/2786-8729.6.2025/332702

STATISTICAL EVALUATION OF PARAMETERS IN NONLINEAR MODELS USING INTEGRAL-FORM OF LEAST SQUARES METHOD AND DIFFERENTIAL NON-TAYLOR TRANSFORMATIONS

Oleksii Pysarchuk
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
https://orcid.org/0000-0001-5271-0248

Oleksandr Tuhanskykh *
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
https://orcid.org/0009-0007-9024-2570

*Corresponding author: ringanaro@gmail.com

The article presents the method for statistical leaning of nonlinear model parameters, its principle and efficiency of application. It proposes a solution to the shortcomings of using the differential spectra balance approach through the integral form of least squares in the scheme of differential non-Taylor transformations.

The object of the study is the process of statistical learning of nonlinear model parameters. The purpose of this paper is to formulate the statistical learning method for evaluation of nonlinear model parameters using LSM in integral form and differential non-Taylor transformations. It is relevant in many areas of modern activity and is necessary for applying the statistical training methodology to a more complex time series, as well as increasing the accuracy of the received expectations.

To achieve this goal the statistical learning methodology was proposed, which is based on the creation of process model in integral form of least squares with simplification using non-Taylor transformations. It differs from existing approaches by incorporating all the available differential discretes in the created model, which allows for better predictability and circumvents the problem of unequal count of discretes inside the models, which allows for better application of the method for different model forms. The algorithm for the process was formed, using which it can be applied different models. In the paper several experiments were conducted to verify the efficacy of the proposed method in different situations. These experiments use generated datasets that are polluted with stochastic errors to better simulate real data.

The results of modeling are shown, and the statistical characteristics of the obtained expectation are compared with the results of the application of the statistical training methodology using the differential spectra balance. During the study, features of the technique were found that must be considered for its application to data with a high number of stochastic deviations. Based on the obtained metrics, a conclusion is made about the effectiveness of using the technique relative to time series reflecting processes of different nature.

Keywords: data science, statistical learning, time series, nonlinear models, differential transformations, least squares method.

1. Introduction

One of the key tasks in the field of Data Science is to smooth the time series to determine the future trend, or to establish individual parameters of the observed process. This is necessary in e-commerce tasks (predicting trading rates, determining retail statistics and patterns in it), as well as in automated control systems (unmanned systems, motion control systems) and other areas of application [3–5]. Time series is a common form of data representation in the following applied areas, and their processing is quite demonstratively represented in the following methods:

– approximations (Moving Average (MA), Exponential Moving Average (EMA), autocorrelation algorithms of the ARIMA type) [2–4, 15];

– statistical learning (smoothing) – Statistical Learning (Least Squares Method (LSM), Kalman filtering) [1–12, 16];

– deep learning using artificial neural networks [5, 8, 11, 17, 18].

Discrete measurements obtained because of observation of a real process always depend on extraneous factors that are nonlinear in nature [2, 3, 9, 10]. This fact causes difficulties in using methods for building mathematical models based on the use of approximation or deep learning. They do not consider a priori known information about the process, and therefore the generated mathematical models [6, 10, 11] have little predictive value and do not allow deriving specific parameters (coefficients) of the process under study. Therefore, the most effective from the point of view of the reliability of the result obtained and its prognostic characteristics is the use of statistical training methods.

As a result of applying such approaches, we most often get a trained model that has a mathematical form, which allows us to generate expectations of the future trend of the time series. But the quality of such a model depends very much on the technique and its use of a priori information about the process. Therefore, it is very urgent to develop new approaches to statistical learning, which will allow considering the largest amount of information about the nonlinear nature of the process.

2. Literature review and problem statement

The task of statistical training of nonlinear models to represent the process under study can be performed by the following well-known approaches:

Smoothing using the polynomial model [1–5] – this approach allows you to build an approximate representation of the process at a certain observation interval, which is a limitation of the approach. Polynomial models do not correspond to the studied process and, accordingly, cannot be used to predict trends with sufficient adequacy [6, 7, 10].

Smoothing of a nonlinear model with linearization to a power polynomial [1, 3–5] – the nonlinear model is formed through decomposition into a Taylor series with the subsequent formation of a power polynomial based on it. This approach allows you to transfer certain information about the process under study to a mathematical model, but the constraints of the convergence interval of the polynomial remain.

Usage of numerical algorithms for iterative search of the model coefficients [2–4, 7] – this approach allows you to obtain the desired parameters of a nonlinear model, without intermediate power polynomials. But the use of numerical algorithms requires the researcher to a priori set initial values to find solutions, the determination of which can be a difficult task for models with many such parameters. Also, depending on the nonlinear complexity of the process, the solution calculation can take a large amount of time, computing power, or simply be impossible due to accumulated errors.

In the papers [6, 7, 12], a new approach to statistical training of the parameters of nonlinear models using a tight approximation of differential spectra [13] through the DSB (Differential Spectra Balance) was proposed. It performs a direct transition between the empirical polynomial model and the theoretical nonlinear one, which allows the determination of unknown parameters of the latter. The analytical description of the process is given a priori depending on the known characteristics of the process under study. This makes the approach effective in solving the problem of building and determining the parameters of models of nonlinear processes. However, this method of statistical training has limitations due to the use of a rigid criterion for approximating theoretical and experimental models [7, 12]. That makes it impossible to apply it to a wide class of models that have an imbalance in the P-spectra of models (have a different number of differential discretes [13]). It also leads to the transfer of both positive and negative characteristics of measurements in the time series (stochastic errors).

The approaches described above allow you to perform statistical training of models, but each of them has its own drawbacks and limitations that make it impossible to apply them to a certain class of problems. This article discusses an alternative approach to statistical training of parameters of nonlinear models, which attempts to solve the main problems of training using DSB.

3. The aim and objectives of the study

The purpose of this paper is to formulate the statistical learning method for evaluation of nonlinear model parameters using least squares method in integral form and differential non-Taylor transformations. It is relevant in many areas of modern activity and is necessary for applying the statistical training methodology to a more complex time series, as well as increasing the accuracy of the received expectations.

To achieve this, the following tasks were set:

1. Form an approach for the training of nonlinear model, using differential non-Taylor transformations through the formation of a mathematical model with the integral form of the LSM. Simplify the resulting integral using differential non-Taylor transformations and for the algorithm for the proposed methodology;

2. Perform experiments with different data and models and compare the results of the proposed approach with previous methods to confirm its efficacy.

4. The study materials and methods of the suggested approach to the learning of nonlinear models using LSM and Differential Non-Taylor transformations

The material given in literature review clearly describes the gap in the ability of modern nonlinear model learning methods. Our previous study [6, 7] suggested an approach that used Differential Spectra Balance to solve this task, but it shows small ability to adapt for more complex scenarios. The main issue is the necessity to have both data polynomial and nonlinear model discretes to be equal in number. That fact often leads to the loss of nonlinear information about the process under study, which leads to lesser prediction results.

To solve this issue, we suggest an approach, when the discretes in question are not solved as a system directly, but firstly summed up in form of residuals. Such a process resembles Least Squares Methods, and so it allows the formation of one singular formula from given discrets. That process formula can be used later to form expressions for the solutions for each coefficient of the original model, effectively giving the solution for the model learning task.

Substituting the balance for the formation of a mathematical model through the integral form of the least squares method, as it presented later, avoids the limitation in the number of discretes. Also, it allows to more accurately describe the nonlinear process, since it can include all possible differential discretes from both empirical and nonlinear models.

At first, we assume that y is a time series obtained through the observation process under study. It has stochastic errors, and is described as:

$$y = \{y_0, y_1, y_2, \ldots, y_n\}, \tag{1}$$

where n is the size of the time series, and y_n is the nth measurement. Approximating function $f(t,c)$ is nonlinear in parameters and was analytically determined from information a priori known about the process under study.

$$f(t,c), c = \{c_i\}, i = 0..m. \tag{2}$$

It is necessary to determine the values of the model coefficients c_i for the smoothing process, the amount of which is m. The selected amount should be satisfactory in terms of the selected metrics to the original time series. We also assume that the polynomial function that approximates the original time series has the form:

$$z(t) = \sum_{i=0}^{m} d_i x^i, \tag{3}$$

where m is rank of the polynomial, which is often equated to the number of parameters of a nonlinear model, and d_i are the coefficients of this polynomial that have been determined using known methods, such as the method of least squares [2, 3].

The method of statistical learning of nonlinear models using DSB, as described in [6, 7, 12], is based on the transfer of characteristics from the experimental data model (2) to a nonlinear model (3). It assumes that the latter is more adequate to the studied process, which is performed using the DNP (Differential Non-Taylor Transformations) method [12]. In this case, the viscous $\varepsilon(t,c)$ between two models is calculated by the formula:

$$\begin{aligned}\delta(c) &= D(\left[P\{z(t)\}_{t^*} \Rightarrow Z(k)\right] - \left[P\{f(t,c)\}_{t^*} \Rightarrow F(k,c)\right] \\ &= D[P\{\varepsilon(t,c)\}_{t^*} \Rightarrow E(k,c)] \rightarrow min\end{aligned} \tag{4}$$

where D is the generalized operations of the selected approximation criterion (hard or soft), and P is the differential transformation operation of the model to its spectrum, described in [12]. The generalized formula for differential transformations is:

$$F(k) = P\{f(t)\}_{t^*} = \frac{H^k}{k!}[\frac{d^k f(t)}{d^t k}]_{t^*} \tag{5}$$

Thus, having performed the transformation (5), the expression (4) is formed. The learning task in this case is reduced to solving a system of equations formed through partial differentiation of the expression by unknown coefficients of the nonlinear model:

$$\frac{\partial\delta(c)}{\partial c} = 0\,. \tag{6}$$

To perform this through a soft approximation criterion, it is necessary to minimize not the residuals between the differential spectra of the two models. Rather, a certain difference between the residuals $\varepsilon(t)$ of these two models from the expression (4) should be minimized. Thus, assuming that at a certain interval *[a, b]* both models are continuous and have an infinite number of derivatives, we can write the expression:

$$\delta(c) = \int_a^b \varepsilon(t)^2 dt = \int_a^b (z(t) - f(t,c))^2 dt \rightarrow min, i = 0..n\,. \tag{7}$$

This expression is the norm of the LSM, and it specifies the operation of transforming the residuals from expression (4) to the integral of the square of the residuals between the experimental and theoretical models. To formulate expressions of unknown coefficients of a nonlinear model, it is necessary to perform a partial differentiation of the expression (7) according to (6):

$$\frac{\partial\delta(c)}{\partial c} = (\frac{\partial\delta(c)}{\partial c_0} = 0, \frac{\partial\delta(c)}{\partial c_1} = 0, ..., \frac{\partial\delta(c)}{\partial c_m} = 0)\,. \tag{8}$$

To simplify the construction of the expression (7), you need to replace the integration operation. This can be done using differential transformations [14]. Considering that both models are continuous, we can write the expression (7) in the form:

$$\delta(c) = H\sum_{k=0}^{m}\left[\left(\frac{t_b}{H}\right)^{k+1} - \left(\frac{t_a}{H}\right)^{k+1}\right] \times \frac{1}{k+1}\sum_{i=0}^{k} E(k-i)E(i) \rightarrow min\,, \tag{9}$$

where m is the upper limit of summation, defined as $m = 2k_{\max} + 1$, $k_{\max}$ is the number of the last non-zero discrete among the experimental and theoretical models.

We assume, that $t_a = 0$ and $t_a = H$, the difference in residual $E(k)$ is the difference between the P-spectra of the experimental and theoretical models. Then, we will be able to obtain the final version of the expression for the formation of a system of equations (8) using the DNP method:

$$\begin{aligned}\delta(c) &= H\sum_{k=0}^{m}\frac{1}{k+1}\sum_{i=0}^{k}(Z(k-i)-F(k-i,c)(Z(i)-F(i,c))\\ &= H\sum_{k=0}^{m}\frac{1}{k+1}\sum_{i=0}^{k}E(k-i,c)E(i,c)=0\end{aligned} \quad . \tag{10}$$

5. Results of research on the suggested methodology for statistical evaluation of nonlinear model parameters

Taking all the above into consideration, it is possible to form a holistic method of statistical training of a nonlinear model, using the integral form of the LSM in the DNP scheme. It can be written in the algorithm form with the following steps:

1. Derive the *P*-spectra of the nonlinear theoretical model (2) and the experimental model (3) using (5) in the form (4).
2. From an image of the model of the process under study in the form of an expression (10).
3. Create and solve a system of equations of the form (8) to determine the unknown parameters of a nonlinear model.

The suggested approach may allow for better predictions from given nonlinear models compared to previous methods. It incorporates mode nonlinear information through the summation of the discrete residuals, which a priori increases its effectiveness. However, it is necessary to test and compare is to the existing approaches.

5.1. Analysis of the effectiveness of the proposed methodology for the exponential model

To check the effectiveness of the statistical training methodology described above, we will conduct a model experiment. To do this, you need to create a data generator from which a time series will be obtained (contaminated with stochastic errors for training, and clean to evaluate the result). It is also necessary to choose an approximation function that will be used as a model for the method described above.

Parameters of the model experiment:

Generating model: $z(t) = d_0 + d_1 t + d_2 t^2 + d_3 t^3$, $coeffs = [0, 0, 0.0005, 0]$

Time series size: 20000

Parameters of stochastic errors: $\sigma = 10000.0$, $coeff = 3.0$, $dens = 10.0$

Approximating model: $f(t,c) = c_0 + c_1 t + c_2 e^{c_3 t}$

Training/Testing intervals: $train = [0,\ 10000]$, $test = [10001,\ 20000]$

According to the algorithm of the method, the first step is to form the *P*-spectra of the generating and approximating functions, respectively. Using the transformation (5), we will form the result in the form (4):

$$\delta(c) = D[P\{\varepsilon(t,c)\}_{t^*} \Rightarrow E(k,c)] = D\begin{pmatrix} d_0 - c_0 - c_2 \\ -H(c_2c_3 + c_1 - d_1) \\ -\dfrac{H^2(c_2c_3^2 - 2d_2)}{2} \\ -\dfrac{H(c_2c_3^3 - 6d_3)}{6} \end{pmatrix}, \tag{11}$$

where c_i are unknown coefficients of the nonlinear approximating function, and – coefficients of the polynomial function of the data that were calculated using the LSM.

The next step is to use the expressions (9) and (10) to form a mathematical model of the underlying process in terms of an integral sum simplified by the image through differential transformations. The upper limit of the summation of m in this case will be 7, since the last discrete number in the spectrum is the third. Therefore, the integral sum will look like:

$$\begin{aligned}\delta(c) &= \frac{H^7\left(c_2c_3^3-6d_3\right)^2}{252}+\frac{H^6\left(c_2c_3^3-6d_3\right)\left(c_2c_3^2-2d_2\right)}{36}\\ &+H^5\left(\frac{\left(c_2c_3^3-6d_3\right)\left(c_2c_3+c_1-d_1\right)}{15}+\frac{\left(c_2c_3^2-2d_2^2\right)}{20}\right)\\ &+H^4\left(-\frac{\left(c_2c_3^3-6d_3\right)\left(d_0-c_0-c_2\right)}{12}+\frac{\left(c_2c_3^2-2d_2\right)\left(c_2c_3+c_1-d_1\right)}{4}\right).\\ &+H^3\frac{\left(c_2c_3+c_1-d_1\right)^2-\left(c_2c_3^2-2d_2\right)\left(d_0-c_0-c_2\right)}{3}\\ &-H^2\left(c_2c_3+c_1-d_1\right)\left(d_0-c_0-c_2\right)+H\left(d_0-c_0-c_2\right)^2\end{aligned} \tag{12}$$

The resulting expression is a generalization of the process under study, and to obtain expressions for the individual coefficients sought, it is necessary to differentiate (12) by unknown parameters of the nonlinear model, forming a system of equations of the form (8):

$$\frac{\partial\delta(c)}{\partial c_0}=H^4\frac{c_2c_3^2-d_3}{12}+H^3\frac{c_2c_3^2-2d_2}{3}+H^2(c_2c_3+c_1-d_1)-2H(d_0-c_0-c_2)=0, \tag{13}$$

$$\begin{aligned}\frac{\partial\delta(c)}{\partial c_1} &= H^5\left(\frac{c_2c_3^3}{15}-\frac{2d_3}{5}\right)+H^4\left(\frac{c_2c_3^2}{4}-\frac{d_2}{2}\right)+H^3\left(\frac{2c_2c_3}{3}+\frac{2c_1}{3}-\frac{2d_1}{3}\right),\\ &-H^2\left(d_0-c_0-c_2\right)=0\end{aligned} \tag{14}$$

$$\begin{aligned}\frac{\partial\delta(c)}{\partial c_2} &= \frac{H^7c_3^3\left(c_2c_3^3-6d_3\right)}{126}+\frac{H^6(c_3^3\left(c_2c_3^2-2d_2\right)+c_3^2\left(c_2c_3^3-6d_3\right))}{36}\\ &+H^5\left(\frac{c_3^3\left(c_2c_3+c_1-d_1\right)}{15}+\frac{c_3\left(c_2c_3^3-6d_3\right)}{15}+\frac{c_3^2\left(c_2c_3^2-2d_2\right)}{10}\right)\\ &+H^4\left(\frac{-c_3^3\left(d_0-c_0-c_2\right)}{12}+\frac{c_2c_3^3}{12}-\frac{d_3}{2}+\frac{a_3^2\left(c_2c_3+c_1-d_1\right)}{4}+\frac{c_3\left(c_2c_3^2-2d_2\right)}{4}\right),\\ &+H^3\left(-c_3^2\frac{d_0-c_0-c_2}{3}+c_2\frac{c_3^2}{3}-\frac{2d_2}{3}+2\frac{c_3\left(c_2c_3+c_1-d_1\right)}{3}\right)\\ &-H^2\left(c_3\left(d_0-c_0-c_2\right)+\left(c_2c_3+c_1-d_1\right)\right)-2H\left(d_0-c_0-c_2\right)=0\end{aligned} \tag{15}$$

$$\frac{\partial\delta(c)}{\partial c_3}=\frac{H^7c_2c_3^2\left(c_2c_3^3-6d_3\right)}{42}+\frac{H^6}{12}c_2c_3^2\left(c_2c_3^2-2d_2\right)$$
$$+\frac{H^6c_2c_3\left(c_2c_3^3-6d_3\right)}{18}+H^5\left(c_2c_3^2\frac{c_2c_3+c_1-d_1}{5}+\frac{c_2\left(c_2c_3^3-6d_3\right)}{15}+\frac{c_2c_3\left(a_2*c_2^2-2d_2\right)}{5}\right). \quad (16)$$
$$+\frac{-c_2c_3^2\left(d_0-c_0-c_2\right)}{4}+c_2c_3\frac{c_2c_3+c_1-d_1}{2}+H^4\frac{c_2\left(c_2c_3^2-2d_2\right)}{4}+\frac{-2c_2c_3\left(d_0-c_0-c_2\right)}{3}$$
$$+2H^3\frac{c_2\left(c_2c_3+c_1-d_1\right)}{3}-c_2H^2\left(d_0-c_0-c_2\right)=0$$

The resulting system of equations can be solved analytically by substituting in (13–16) the calculated values of the coefficients d_i. As a result, we will get a matrix, which will be a set of possible coefficients of a nonlinear model $f(t,c)$:

1. $c_0=-0.00008366699999, c_1=0.0005010000000, c_2=0, c_3=-1.996007988;$

2. c_0=-0.00008366699999 -$c_2, c1=0.0005010000000, c2=c2, c3=0;$

3. $c_0=-0.00001859265000, c_1=0.0002087498210, c_2=1.531289395\times10^{-22},$
$c_3=2.331435453\times10^{6};$

4. $c_0=-4.048433295\times10^{9}, c_1=2012.072142, c_2=4.048433295\times10^{9},$
$c_3=-4.970001961\times10^{-7}.$

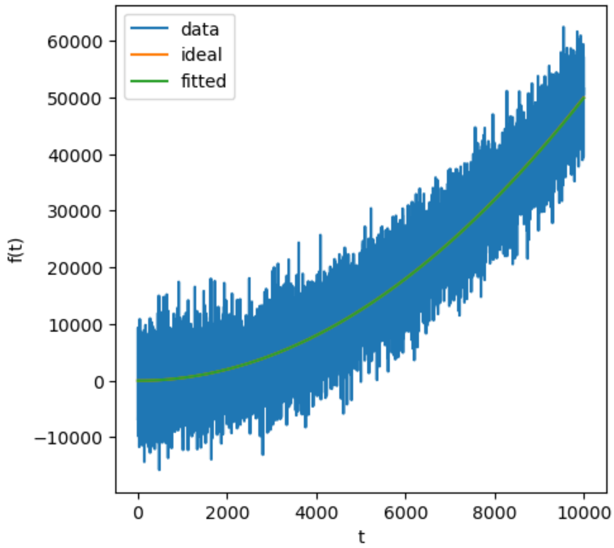
From the above list, some of the decisions are not adequate. Thus, solution **No. 1** has among the solutions found $c_2=0$, which completely cancels out the exponential part of the nonlinear model. And solution **No. 2** generally has $c_2=c_2$, as well as zeros and unresolved fractions. Thus, when choosing the final solution, it is necessary to filter out the roots of the system of equations, referring to their adequacy, and quality relative to the selected metrics for smoothing. So, when choosing between solutions **No. 3** and **No. 4**, if you check them, it is **No. 4** that gives the best indicators, so we will use it as the result of the learning problem.

To verify the result, we will run a simulation of the model with the found parameters, at the intervals of training and testing (prediction). We follow its metrics, such as Residual Squared Error (RSE), Standard Error (*StdErr*), Determination Coefficient (*Rsq*), and Concordance Coefficient (*CCord*). The results of said simulation are shown in Figure 1 and Table 1. Figure 1 shows as horizontal axis the time value (marked as t) for the time series, and vertical axis (marked as $f(t)$) is the value of said time series for that time value.

To compare the obtained metrics, we will use the results of smoothing the same time series using a technique that uses a rigid approximation criterion and was presented in the papers [6, 7].

From the data obtained, it can be concluded that although both techniques adequately fulfilled the learning task at both intervals. It was the variation with a soft approximation criterion (integral form) that showed better results in terms of *RSE*. At the training interval, DSBI (Differential Spectra Balance Integral) has an advantage of ~60% in *RSE* over regular DSB (Differential Spectra Balance). In the testing interval, the discrepancy increases to ~140%, which indicates a much higher quality of smoothing precisely because of DSBI. Both methods now show an adequate result in terms of *Rsq* and *CCord* coefficients.

However, it should be noted that the method of statistical training using the integral form of the LSM has certain drawbacks due to the formation of the integral sum. In such a situation, even small errors in the initial data due to the presence of the sum can lead to significant deviations in the results obtained.



a *b*

Fig. 1. Visualization of the results of modeling a learned exponential function at the intervals: *a* – training, *b* – testing.

Table 1. Metrics of smoothing results and time series prediction using an exponential model

		RSE	***StdErr***	***Rsq***	***CCord***
train	**dsb**	50.457	148.677	1.000	1.000
	dsbi	31.264	148.826	1.000	1.000
test	**dsb**	847.731	440.390	1.000	1.000
	dsbi	352.039	432.928	1.000	1.000

Thus, during the experiment, it was found that the low-impact coefficients, which are left among the solutions by the training of the polynomial experimental model according to the usual least squares. It led to critical errors or made it impossible to solve the formed system of equations of the form (8). That is, because of training the experimental model according to the LSM in this experiment, the following was obtained:

1. $d_0 = 1.43803739 \times 10^{-14}$;

2. $d_1 = -9.97049346 \times 10^{-17}$;

3. $d_2 = 5.00000000 \times 10^{-4}$;

4. $d_3 = 1.77882991 \times 10^{-22}$.

Obviously, the coefficients d_0, d_1, d_3 here are the remnants of the LSM itself, because they are too small to significantly affect the result. But when using DSBI, due to the presence of an integral sum, such values give strong deviations, which interferes with the effective application of the technique.

To solve this problem, it is proposed to apply a filter over the results obtained by the LSM, zeroing out all coefficients that will be too small to be accepted for further processing. In such a filter, you need to consider the order of the coefficient itself, since in a polynomial the larger it is, the more influential the parameter itself becomes. It is necessary to base the criterion on the value of the stochastic error that is allowed for the coefficient of the corresponding order and cancel all values that are less than such a criterion. Thus, based on the principles described above, it is possible to form a criterion:

$$filter(d, i) = log(d_i, 10) + 3i - 1 \quad (17)$$

where d are the coefficients of the polynomial that need to be filtered, and i is the order of the next coefficient. If such a filter returns to a positive value, it means that the coefficient is greater than the expected stochastic error (10^{-3} of rank) and accordingly must be taken for further processing. If the filter returns to a negative value, then the resulting coefficient is considered less than the error, and, accordingly, must be canceled.

By applying criterion (17) during training according to the LSM of the experimental model, the result was obtained: $d_0 = 0.0, d_1 = 0.0, d_2 = 5.00000000 \times 10^{-4}, d_3 = 0.0$

Using such filtering of the parameters of the polynomial experimental model, it was possible to achieve the smoothing indicators for DSBI, which are demonstrated in Figure 1 and in Table 1.

5.2. Analysis of the proposed methodology for the transcendental model.

To check the effectiveness of using the statistical learning methodology with the integral form of LSM, another experiment was conducted, in which an analytical transcendental model was applied. Such a check will show the operation of the method in a more complex situation, when the dependencies between the internal parameters of the model are nonlinearly related.

Parameters of the model experiment:

Generating function: $z(t) = d_0 + d_1 t + d_2 t^2 + d_3 t^3,\ coeffs = \left[4785.5, -4.9093, -0.0034, 10^{-6}\right]$

Time series size: 2000

Parameters of stochastic errors: $\sigma = 1000.0, coeff = 3.0, dens = 10.0$

Approximating function: $f(t,c) = c_0 cos(c_1 t) + c_2 sin(c_1 t)$

Training/Testing intervals: $train = [0,\ 1000],\ test = [1001,\ 2000]$

In the same way as for the exponential model, we apply the algorithm of the technique, and form an expression of the residual between the P-spectra of the models in the form (4):

$$\delta(c) = D\begin{pmatrix} d_0 - c_0 \\ -H(c_1 c_2 - d_1) \\ \dfrac{H^2(c_0 c_1^2 - 2d_2)}{2} \\ d_3 H^3 \end{pmatrix} \tag{18}$$

Next, we form an integral sum (9) through differential transformations of the expression (10), and take the indicator $m = 7$. Since the maximum discrete number in the experimental model is tree. The resulting sum for the expression (18):

$$\begin{aligned} \delta(c) = {} & \frac{d_3^2 H^7}{7} + \frac{H^6 d_3 (c_0 c_1^2 + 2d_2)}{6} + H^5\left(\frac{-2d_3(c_1 c_2 - d_1)}{5} + \frac{(c_0 c_1^2 + 2d_2)^2}{20}\right) \\ & + H^4\left(\frac{d_3(d_0 - c_0)}{2} - \frac{(c_0 c_1^2 + 2d_2)(c_1 c_2 - d_1)}{4}\right) \\ & + H^3\left(\frac{(c_0 c_1^2 + 2d_2)(d_0 - c_0)}{3} + \frac{(c_1 c_2 - d_1)^2}{3}\right) \\ & - H^2(c_1 c_2 - d_1)(d_0 - c_0) + H(d_0 - c_0)^2 \end{aligned} \tag{19}$$

The next step is to differentiate (16) by the unknown parameters of the nonlinear model, according to (8). In this case, these will be the coefficients c_i when the parameters of the polynomial model d_i are known and calculated through LSM relative to the initial time series. The resulting expressions are of the form (17):

$$\frac{\partial\delta(c)}{\partial c_0}=\frac{H^6 d_3 c_1^2}{6}+\frac{H^5 c2_1\left(c_0c_1^2+2d_2\right)}{10}+H^4\left(\frac{-d_3}{2}-\frac{c^2_{(c_1c_2-d_1)}}{4}\right)$$
$$+H^3\left(\frac{c_1^2\left(d_0-c_0\right)}{3}-\frac{c_0c_1^2}{3}-\frac{2d_2}{3}\right) \tag{20}$$
$$+H^2\left(c_1c_2-d_1\right)-2H\left(d_0-c_0\right)=0$$

$$\frac{\partial\delta(c)}{\partial c_1}=\frac{H^6 d_3 c_0 c_1}{3}+H^5\left(-\frac{2d_3c_2}{5}+\frac{c_0c_1\left(c_0c_1^2+2d_2\right)}{5}\right)$$
$$+H^4\left(-\frac{c_0c_1\left(c_1c_2-d_1\right)}{2}-\frac{c_2\left(c_0c_1^2+2d_2\right)}{4}\right) \tag{21}$$
$$+H^3\left(\frac{2c_0c_1\left(d_0-c_0\right)}{3}+\frac{2c_2\left(c_1c_2-d_1\right)}{3}\right)-c_2H^2\left(d_0-c_0\right)=0$$

$$\frac{\partial\delta(c)}{\partial c_2}=-\frac{H^5 2d_3c_1}{5}-\frac{H^4c_1\left(c_0c_1^2+2d_2\right)}{4}+\frac{2H^3c_1\left(c_1c_2-d_1\right)}{3} \tag{22}$$
$$-H^2c_1\left(d_0-c_0\right)=0$$

We solve the obtained system of equations (20–22) with respect to the unknown coefficients of the nonlinear model, substituting the parameters of the experimental polynomial model found by the LSM:

1. c_0=4783.039303, c_0=0, $c2=0$;

2. c_0=4785.499998, c_1=0.001192750974, c_2=-4115.942072;

3. c_0=4785.499998, c_1=-0.001192750974, c_2=4115.942072

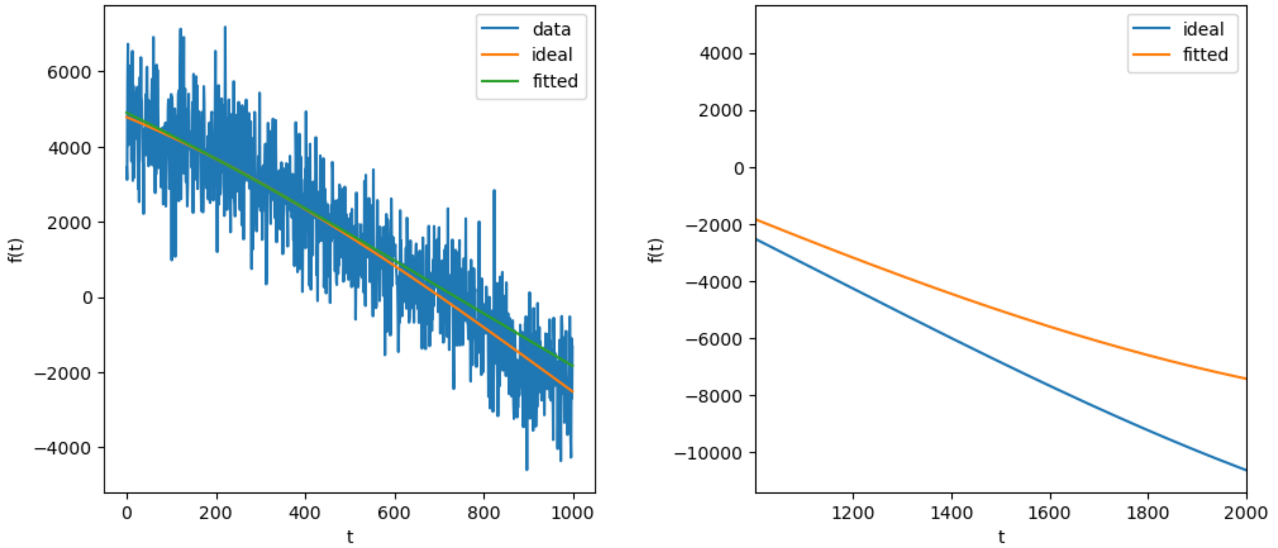
4. $c_0=0, c_1=3.457187076I, c_2=-2070.761617I$

5. $c_0=0, c_1=-3.457187076I, c_2=2070.761617I$

Among the solutions obtained, we can again see a set of inadequate ones. Thus, **No. 1, 4, 5** have zero values of coefficients, and the last two generally have imaginary values, that is, complex roots were obtained.

The obtained DSB or DSBI smoothing results are often used further in numerical algorithms to optimize the approximation result to improve performance. Such methods are numerous, and, accordingly, cannot use complex solutions. Therefore, all solutions that will have imaginary particles must also be discarded.

Thus, solutions **No. 2, 3** remains, which are symmetrical and differ only in the position of the negative sign. Since we know that we are teaching the transcendental model, due to its periodic nature, it can be concluded that these solutions are equivalent. We chose one of them to run a simulation to check the quality of the result. The results of the simulation are shown in Figure 2 and Table 2. Figure 2*a*, *b* shows as horizontal axis the time value value (marked as *t*) for the time series, and vertical axis (marked as *f(t)*) is the value of said time series for that time value.



a *b*

Fig. 2. Visualization of the results of modeling the learned transcendental model at the intervals: *a* – training, *b* – testing.

As in the case of the exponential model, we will use a training methodology with a rigid approximation criterion to compare the effectiveness of the obtained metrics.

As a result of modeling, both methods obtained an adequate result at the training interval, and their indicators are very similar to each other. At the testing interval, both methods have a significant deterioration in metrics, which indicates a strong influence of nonlinear complexity on the result.

Overall, both DSB and DSBI showed a similar result in the transcendental model training, ~0.99 of concordance coefficient and training phase and ~0.70 in the testing phase. The coefficient of determination in both cases is negative, indicating that both models will further diverge from the ideal trend, and a higher *RSE* in DSBI indicates a greater rate of divergence in it.

Table 2. Metrics of smoothing results and time series prediction according to the transcendental model

		RSE	***StdErr***	***Rsq***	***CCord***
train	**dsb**	222.853	63.250	0.988	0.996
	dsbi	264.732	62.240	0.982	0.995
test	**dsb**	1819.289	53.698	-0.147	0.715
	dsbi	1998.442	51.717	-0.492	0.670

Low scores in both methods indicate a common problem between them, namely a low amount of nonlinear process information displayed in the time series. As can be seen in Figure 2, it is a simple arc, although a priori it is known about the periodicity of the process. This may explain why statistical learning techniques have difficlty establishing the true form of the process under study.

6. Analysis of the results obtained during the research on the methodology for statistical evaluation of nonlinear model parameters

As a result of the research, the effectiveness of the application of the method of statistical training of nonlinear models using the integral form of the LSM and differential non-Taylor transformations was shown. Among the experiments carried out, this approach showed itself best in working with the exponential model, and with the periodic model it has indicators on a par with its analogues. In both cases, the application of the proposed technique led to an improvement in expectation metrics, which proves its effectiveness.

During the experiments, certain conditions were also established that must be met for the effective application of the proposed statistical training methodology:

Due to the use of the integral form of the LSM, small errors in the preliminary calculations of the coefficients of the experimental model led to a large deviation of the result. It is proposed to filter the calculated coefficients of the experimental model using (14) to reduce the impact of possible residuals from the training of the polynomial model.

As a result of solving a system of equations of the form (8), inadequate solutions often appear, which must be filtered by certain features. It is proposed to use to search for correct solutions: the absence of zero coefficients, the absence of imaginary fractions in solutions, the best metrics of the resulting trained model.

A strong nonlinear relationship between the parameters of the model affects the effectiveness of the technique. This can be solved by extending the time series to include more nonlinear information about the process under study and, accordingly, it is better possible to perform model training.

The methodology considered has shown its effectiveness and potential for further optimization of relatively different forms of models and time series. The issue of adaptation of the method to complex nonlinear processes and the possibility of its application in relation to other similar approaches requires additional research.

Conclusions

The study described the solution of the relevant scientific task of learning the nonlinear model parameters using LSM in integral form to create the process model. The research has several results:

Proposed and described the methodology for the training of nonlinear models which incorporates all available differential discretes. It also uses the use of non-Taylor differential transformations to simplify and solve the integral expression of the process under study.

The model experiments were conducted to evaluate the methodology application results, which were compared to the existing statistical learning method. Proposed methodology showed effectiveness in both experimental scenarios and outperformed the other method.

Using the proposed approach for statistical evaluation of the parameters of nonlinear models it is possible to increase the predictive value of existing nonlinear models. It includes more nonlinear information about the process during training which both removes the disbalance in differential discretes problem and improves the predictability.

The proposed methodology was analyzed, and some specifics of its application were found, which suggests that more research is needed. The methodology can be improved and combined with existing approaches to produce better autonomous results.

References

[1] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, An Introduction to Statistical Learning. Springer, 2023, https://doi.org/10.1007/978-3-031-38747-0.

[2] M. Nguyen, A Guide on Data Analysis. Bookdown, 2020. [Online] Available: https://bookdown.org/mike/data_analysis/

[3] S. C. Chapra and R. P. Canale, Numerical Methods for Engineers, 7th ed. New York, NY, USA: McGraw-Hill Education, 2015.

[4] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. New York, NY: Springer New York, 2009, https://doi.org/10.1007/978-0-387-84858-7.

[5] F. Provost and T. Fawcett, Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking. Beijing, China: O’Reilly Media, 2013.

[6] Oleksandr Tuhanskykp, D. Baran, and Oleksii Pysarchuk, “Method for Statistical Evaluation of Nonlinear Model Parameters in Statistical Learning Algorithms,” Lecture notes in networks and systems, pp. 265–274, Jan. 2024, https://doi.org/10.1007/978-981-97-3559-4_21.

[7] O. O. Pysarchuk, O. A. Tuhanskykh, and D. R. Baran, “Method for determining the structure of nonlinear models for time series processing” Radio Electronics Computer Science Control, no. 1, pp. 53–62, Apr. 2025, https://doi.org/10.15588/1607-3274-2025-1-5.

[8] D. Dietrich, B. Heller, and B. Yang, Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. Indianapolis, IN, USA: Wiley, 2021.

[9] S. Konishi, Introduction to Multivariate Analysis. Informa, 2014, https://doi.org/10.1201/b17077.

[10] E. W. Grafarend, S. Zwanzig, and J. L. Awange, Applications of Linear and Nonlinear Models. Cham: Springer International Publishing, 2022, https://doi.org/10.1007/978-3-030-94598-5

[11] A. W. Roddam, “Measurement Error in Nonlinear Models: A Modern Perspective,” Journal of the Royal Statistical Society Series A (Statistics in Society), vol. 171, no. 2, pp. 505–506, Mar. 2008, https://doi.org/10.1111/j.1467-985x.2007.00528_4.x.

[12] V. P. Kharchenko, O. O. Pysarchuk Nelinijne ta bagatokryterialne modelyuvannya procesiv u systemax keruvannya ruxom, Kyiv, Ukraine: Instytut obdarovanoyi dytyny, 2015.

[13] G. E. Puhov, Dyferentsiini spektry ta modeli. Kyiv, Ukraine: Naukova Dumka, 1990.

[14] G. E. Puhov, Dyferentsiini peretvorennia ta matematychne modeliuvannia fizychnykh protsesiv. Kyiv, Ukraine: Naukova Dumka, 1986.

[15] R. H. Shumway and D. S. Stoffer, Time Series Analysis and Its Applications. New York, NY: Springer New York, 2011, https://doi.org/10.1007/978-1-4419-7865-3.

[16] M. K. Transtrum and J. P. Sethna, “Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization,” arXiv (Cornell University), Jan. 2012, https://doi.org/10.48550/arxiv.1201.5885.

[17] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, “TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis,” Oct. 2022, https://doi.org/10.48550/arxiv.2210.02186.

[18] M. Jin et al., “Time-LLM: Time Series Forecasting by Reprogramming Large Language Models,” arXiv (Cornell University), Oct. 2023, https://doi.org/10.48550/arxiv.2310.01728.

УДК 519.2+004.8

СТАТИСТИЧНЕ НАВЧАННЯ ПАРАМЕТРІВ НЕЛІНІЙНИХ МОДЕЛЕЙ З ІНТЕГРАЛЬНОЮ ФОРМОЮ МЕТОДУ НАЙМЕНШИХ КВАДРАТІВ В СХЕМІ ДИФЕРЕНЦІЙНО-НЕТЕЙЛОРІВСЬКИХ ПЕРЕТВОРЕНЬ

Олексій Писарчук
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна
https://orcid.org/0000-0001-5271-0248

Олександр Туганських
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна
https://orcid.org/0009-0007-9024-2570

У статті представлено метод статистичного навчання параметрів нелінійної моделі, його принцип та ефективність застосування. Запропоновано вирішення недоліків використання підходу балансу диференційних спектрів через інтегральну форму МНК у схемі диференціальних нетейлорових перетворень.

Об'єктом дослідження є процес статистичного навчання параметрів нелінійної моделі. Метою даної роботи є визначення та формалізація статистичного методу навчання для оцінювання параметрів нелінійної моделі з використанням МНК в інтегральній формі та диференціальних нетейлорових перетворень. Вона актуальна в багатьох сферах сучасної діяльності і необхідна для застосування методики статистичного навчання до більш складних часових рядів, а також підвищення точності отриманих очікувань.

Для досягнення поставленої мети була запропонована методика статистичного навчання, яка базується на створенні моделі процесу в інтегральній формі МНК зі спрощенням за допомогою диференційних нетейлорівських перетворень. Він відрізняється від існуючих підходів тим, що в створену модель включені всі доступні диференціальні дискрети, що дозволяє досягти кращої передбачуваності та обходити проблему нерівного числа дискрет всередині моделей, що дозволяє краще застосовувати метод для моделей різних форм. Був сформований алгоритм виконання процесу, за допомогою якого можна застосовувати різні моделі. У роботі було проведено кілька експериментів для перевірки ефективності запропонованого методу в різних ситуаціях. У цих експериментах використовуються згенеровані набори даних, забруднені стохастичними помилками, для кращої симуляції реальних даних.

Показано результати моделювання та проведено порівняння статистичних характеристик отриманого очікування із результатами застосування методики статистичного навчання з використанням балансу диференційних спектрів. Під час дослідження були віднайдені особливості методики, які необхідно приймати до уваги для її застосування до даних з високою кількістю стохастичних відхилень. На основі отриманих метрик зроблено висновок про ефективність використання методики відносно часових рядів, що відображають процеси різної природи.

Ключові слова: аналіз даних, статистичне навчання, часові ряди, нелінійні моделі, диференційні перетворення, метод найменших квадратів.