PHYSICS AND MATHS

Gorodetskyi Viktor
PhD, associate professor, National Technical University of Ukraine
«Igor Sikorsky Kyiv Polytechnic Institute», Ukraine

Osadchuk Mykola
PhD, assistant professor, National Technical University of Ukraine
«Igor Sikorsky Kyiv Polytechnic Institute», Ukraine

ON IDENTIFICATION OF DYNAMICAL SYSTEM USING SCALAR TIME SERIES

Abstract. *We propose an algorithm for reconstruction from time series a mathematical model in the form of a system of ordinary differential equations. The method is based on analytical transformations connecting the sought-for original system and a system of a special type. The latter has a fractional rational function only in one of the equations.*

Keywords*: ordinary differential equations, original system, standard system, time series*

Introduction

Sometimes, the only information about a process under study can be a discrete time series of data on one of the variables of this process. To obtain a mathematical model in this case, it is necessary to solve the problem of reconstructing the system from a single observable variable [1, 2]. Let the system of ordinary differential equations

$$\begin{cases} \dot{x}_1 = X_1(x_1,...,x_n), \\ \cdots \\ \dot{x}_n = X_n(x_1,...,x_n), \end{cases} \tag{1}$$

describes some process, where $x_1,...,x_n$ *- process variables;* $X_1,...,X_n$ *- continuous (generally speaking, nonlinear) functions. System (1) will be called [3] the original system (OS).*

Let the only observable variable is. A widespread approach [4, 5], in which the unknown variables of system (1) are replaced by time derivatives, and instead of (1), system (2) is considered

$$\begin{cases} \dot{y}_1 = y_2, \\ \dot{y}_2 = y_3, \\ \dots \\ \dot{y}_n = Y(y_1, \dots, y_n), \end{cases} \tag{2}$$

where Y *is some function that is chosen so that the solution of system (2) (function* $y_1(t)$*) coincides with the observed OS variable. Moreover, in some cases [5] this coincidence may be exact (*$x_1(t) \equiv y_1(t)$*), while in others [6, 7] we are talking about a more or less accurate approximation over a certain period of time (*$x_1(t) \approx y_1(t)$*).*

System (2) is called the standard system (SS), and the function Y is called the standard function (SF) [8]. Polynomials of various types [6], functions of the form "polynomial plus fraction" [3, 5], and the ratio of polynomials [5, 9]:

$$Y(y_1, ..., y_n) = \frac{P_1(y_1, ..., y_n)}{P_2(y_1, ..., y_n)}. \tag{3}$$

Obviously, case (3) is the most universal of the three listed, since the other two types of SF are easily reduced to (3).

The purpose of this work is to improve the algorithm proposed in [5] for determining the coefficients of the standard system to improve the accuracy of the results.

1. Basics of the approach

As shown in [5], when using the SF of the form (3), one can construct a simple algorithm for calculating its coefficients. Let in system (2) $n = 3$. Then SS (2) will have the form:

$$\begin{cases} \dot{y}_1 = y_2, \\ \dot{y}_2 = y_3, \\ \dot{y}_3 = Y(y_1, y_2, y_3), \end{cases} \tag{4}$$

where

$$Y(y_1, y_2, y_3) = \frac{\sum_i N_i p_i(y_1, y_2, y_3)}{\sum_i D_i p_i(y_1, y_2, y_3)}. \qquad (5)$$

In expression (5), the numerator and denominator are polynomials, N_i и D_i are constant coefficients, $p_i\left(y_1, y_2, y_3\right)$ are the products of degrees of the SS variables. In this article, we used the functions shown in Table 1, which correspond to the polynomial of the third degree. If necessary, Table 1 can be supplemented with products of higher degrees.

Table 1

Products of degrees of the SS variables in (5)

i	$p_i\left(y_1, y_2, y_3\right)$	i	$p_i\left(y_1, y_2, y_3\right)$	i	$p_i\left(y_1, y_2, y_3\right)$	i	$p_i\left(y_1, y_2, y_3\right)$
0	1	5	$y_1 y_2$	10	y_1^3	15	$y_1 y_3^2$
1	y_1	6	$y_1 y_3$	11	$y_1^2 y_2$	16	y_2^3
2	y_2	7	y_2^2	12	$y_1^2 y_3$	17	$y_2^2 y_3$
3	y_3	8	$y_2 y_3$	13	$y_1 y_2^2$	18	$y_2 y_3^2$
4	y_1^2	9	y_3^2	14	$y_1 y_2 y_3$	19	y_3^3

Note that multiplying the numerator and denominator (5) by the same constant does not change the value of $Y(y_1, y_2, y_3)$. Therefore, the values of the coefficients (5) cannot be calculated uniquely. To eliminate this situation, one can fix one of the coefficients, considering it known. For example, in [5] it was accepted $D_0 = 1$. Then, after reducing to a common denominator and grouping the terms, expression (5), taking into account the table1 will take the form:

$$\begin{aligned} &N_0 + N_1 y_1 + N_2 y_2 + \ldots + N_{18} y_2 y_3^2 + N_{19} y_3^3 - D_1 y_1 \dot{y}_3 + \\ &+D_2 y_2 \dot{y}_3 - \ldots - D_{18} y_2 y_3^2 \dot{y}_3 - D_{19} y_3^3 \dot{y}_3 = D_0 \dot{y}_3, \end{aligned} \qquad (6)$$

where $D_0 = 1$.

Since the values of y_1, $y_2 = \dot{y}_1$, $y_3 = \dot{y}_2$, $\dot{y}_3$ are easily calculated based on the data on the observed time series, it is easy to obtain an algebraic system of equations in which the coefficients N_i and D_i of equation (6) become unknown. When using this equation to form an algebraic system, a situation may occur when $D_0 = 0$. Then, having previously accepted, we will receive an erroneous result when calculating the coefficients. Therefore, it is necessary to be able to fix any of the SF coefficients. In addition, as will be shown below, for real systems it can be known in advance which SF coefficients are equal to zero. Therefore, the proposed algorithm provides for the possibility of assigning an arbitrary value to any of the SF coefficients. In this case, it is advisable to use a more general form of equation (6):

$$\sum_i N_i p_i(y_1, y_2, y_3) - \dot{y}_3 \sum_i D_i p_i(y_1, y_2, y_3) = \\ = \dot{y}_3 \sum_i \tilde{D}_i p_i(y_1, y_2, y_3) - \sum_i \tilde{N}_i p_i(y_1, y_2, y_3), \tag{7}$$

where the terms are grouped on the left, in which N_i *and* D_i *are unknown coefficients, and on the right -* $\tilde{N}_i$, $\tilde{D}_i$ *are the SF coefficients, the values of which are assumed to be known.*

The number of points N_e in the time series, which is sufficient to construct such an algebraic system, is equal to the number of equations of the system, which, in turn, is equal to the number of unknown coefficients of equation (6). In practice, it is advisable to form not one, but many systems of algebraic equations using different sets of points in the time series. Then, having received a larger number of systems for finding the coefficients, the results of solving all systems can be averaged. This approach leads to an increase in the accuracy of determining the SS coefficients.

2. Construction of the algorithm

Based on the above, the following algorithm for calculating the SS coefficients can be proposed. Let $y_1(j)$ be a discrete time sequence of values of the observed OS variable with a sampling step Δt and containing N points. Using the equations of system (4), it is possible to form discrete time sequences $y_2(j),\ y_3(j),\ \dot{y}_3(j)$ using numerical differentiation. To obtain a system of algebraic equations that allows calculating the SF coefficients (5), a vector time sequence is formed

$$v(j) = \{y_1(j),\ y_2(j),\ y_3(j),\ \dot{y}_3(j)\}, \qquad (8)$$

Taking into account the general form of the polynomials of the numerator and denominator, as well as the number of known coefficients, the number of unknown SS coefficients N_e is determined. Further, for a certain point of the time series with a number k, vectors $v(k)$, $v(k+\Delta n)$, $v(k+2\Delta n)$,…, $v(k+(N_e-1)\Delta n)$ are formed, where Δn is the number of points in the time series, which specifies the interval for the formation of the algebraic system. The known values of the components of each generated vector, as well as the known SS coefficients $\tilde{N}_i$, $\tilde{D}_i$, are substituted in (7). As a result, a system of N_e linear equations is obtained, which is solved with respect to the SS coefficients. The position of the starting point changes within the limits $k = 0,...,N_s$, where

$$N_s = N - (N_e - 1)\Delta n - 1. \qquad (9)$$

For each k value, a system of linear equations is compiled and solved, the solutions of which are the SS coefficients $N_i(k)$, $D_i(k)$. For each SS coefficient, its mean value and standard deviation are calculated:

$$\bar{X} = \frac{1}{N_s}\sum_{k=0}^{N_s-1} X(k), \qquad (10)$$

$$\sigma(X) = \sqrt{\frac{1}{N_s}\sum_{k=0}^{N_s-1}(X(k) - \bar{X})^2}, \qquad (11)$$

where $X(k) \equiv N_i(k)$ *or* $X(k) \equiv D_i(k)$ *depending on whether the coefficients refer to the numerator or denominator of (5). It can be assumed that at some moments of time the denominator of the SF (5) may tend to zero, and the values of the solutions of the system of linear equations may increase by several orders of magnitude, which will introduce an error when averaging according to formula (10). To eliminate this* *error, the obtained values of the coefficients* $X(k)$ *are filtered in accordance with the condition*

$$\left|X(k)-\bar{X}\right| \le K_s \sigma(X), \tag{12}$$

where K_s *is the filtering parameter.*

If at a point k condition (12) is not satisfied for at least one SS coefficient, then all solutions of the system of linear equations at this point are excluded from further consideration. After checking condition (12) for all N_s solutions and removing solutions for which the condition is not satisfied, N_s decreases, values of (10) and (11) are recalculated, and the solutions are filtered again. The filtering is repeated as long as condition (12) is violated at least once. In other words, after filtering, only those solutions remain whose value differs from the mean by no more than K_s values of standard deviations. If necessary, the filtration step can be skipped.

To get solutions of systems of linear equations, average values are calculated according to (10), which are taken as the coefficients of the SS. Next, the correctness of determining the SS coefficients is checked by numerical integration of SS by the 4th order Runge-Kutta method. The vector of values of the observed variable and its derivatives at the initial point of the time sequence $v(0)$ is taken as the initial conditions. The result of integration is a reconstructed time series $y_1'(j)$. Integration continues until the length of the reconstructed time series N_r becomes equal to the length of the investigated time series N, or until the condition

$$\left|y_1'(j)-y_1(j)\right| \le 10^4 \tag{13}$$

is violated which means the divergence of the computational process.

If the SF has the form (5), then a singularity may appear when the denominator of the SF becomes zero, which can lead to an integration error. Elimination of the error is carried out as follows: if between the points with numbers j and $j+1$ the denominator of the SF changes sign, then the result of the $(j+1)$th step of the Runge-Kutta method is discarded and $v(j+1)$ accepted instead.

Next, an assessment of the difference between the investigated and reconstructed time sequences is carried out. To do this, one can use the root-mean-square deviation

$$K_1 = \sqrt{\frac{1}{N_r}\sum_{j=0}^{N_r-1}(y_1'(j) - y_1(j))^2}\,, \qquad (14)$$

smaller values of which correspond to more accurate reconstruction. In order to compare the standard deviations calculated for different initial time sequences, regardless of the amplitude and the number of points, division (14) by the mean square of the original time sequence is performed.

$$K_2 = \frac{\sqrt{\frac{1}{N_r}\sum_{j=0}^{N_r-1}(y_1'(j) - y_1(j))^2}}{\sqrt{\frac{1}{N}\sum_{j=0}^{N-1}y_1(j)^2}}\,. \qquad (15)$$

To estimate the length of the reconstructed time sequence, we used the value

$$K_3 = N - N_r\,. \qquad (16)$$

Conditions (15) and (16) were combined to calculate the "quality" index of the reconstruction:

$$K = \frac{\sqrt{\frac{1}{N_r}\sum_{j=0}^{N_r-1}(y_1'(j) - y_1(j))^2}}{\sqrt{\frac{1}{N}\sum_{j=0}^{N-1}y_1(j)^2}} + P(N - N_r)\,, \qquad (17)$$

where P *is a positive penalty coefficient. The value of the latter is chosen in such a way that, for* $N_r < N$*, the value of (17) will be obviously greater (worse) than for any case with* $N_r = N$*. In this study, value* $P = 10$ *was taken.*

The algorithm provides for varying the step Δn of selecting points for systems of linear equations to find a step for which the reconstruction quality index is of the lowest value. The values of the SS coefficients corresponding to the minimum index of the reconstruction quality (17) are taken as the result.

3. Results

The proposed algorithm was applied to the modeling of OSs related to the R-class [8], i.e. systems like

$$\begin{cases} \dot{x}_1 = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3, \\ \dot{x}_2 = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3, \\ \dot{x}_3 = c_0 + c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_1 x_2 + c_5 x_1 x_3 + c_6 x_2 x_3. \end{cases} \quad (18)$$

The SS corresponding to OS (18) has the form

$$\begin{cases} \dot{y}_1 = y_2, \\ \dot{y}_2 = y_3, \\ \dot{y}_3 = \dfrac{1}{D_0 + D_1 y_1 + D_2 y_2}\left(N_0 + N_1 y_1 + N_2 y_2 + N_3 y_3 + N_4 y_1^2 + N_5 y_1 y_2 + \right. \\ + N_6 y_1 y_3 + N_7 y_2^2 + N_8 y_2 y_3 + N_9 y_3^2 + N_{10} y_1^3 + N_{11} y_1^2 y_2 + N_{12} y_1^2 y_3 + \\ \left. + N_{13} y_1 y_2^2 + N_{14} y_1 y_2 y_3 + N_{15} y_1 y_3^2 + N_{16} y_2^3 \right). \end{cases} \quad (19)$$

The R-class includes a number of chaotic systems, including the Rössler system [10], which, taking into account the notation (18), has the form:

$$\begin{cases} \dot{x}_1 = a_2 x_2 + a_3 x_3, \\ \dot{x}_2 = b_1 x_1 + b_2 x_2, \\ \dot{x}_3 = c_0 + c_3 x_3 + c_5 x_1 x_3. \end{cases} \quad (20)$$

This system was solved by the 4th order Runge-Kutta method with coefficients $a_2 = a_3 = -1$, $b_1 = c_5 = 1$, $b_2 = 0.15$, $c_0 = 0.2$, $c_3 = -10$ on a time interval of 40 s with a step of 0.002 s. Time series and phase portraits of system (20) are shown in Fig. 1. A variable x_1 was taken as an observable. Using the proposed algorithm, the values of the SS coefficients (19) were determined for three cases: without filtering the SS coefficients, filtering with $K_S = 6$ and filtering with $K_S = 3$ according to (12). For each case, the value Δn was varied within $1..1000$ points and the value Δn was chosen at which the quality index was the smallest. It was found that in the absence of filtration $\Delta n = 520$ and $\mathrm{K} = 0.00495$; at $K_S = 6$ - $\Delta n = 360$ and $\mathrm{K} = 0.00072$; at

$K_S = 3$ - $\Delta n = 359$ and $\mathrm{K} = 0.00097$. The average values of the SS coefficients and standard deviations are given in Table. 2. The value of the coefficient D_0 was accepted as known, therefore in Table 2 its standard deviation is equal to zero.

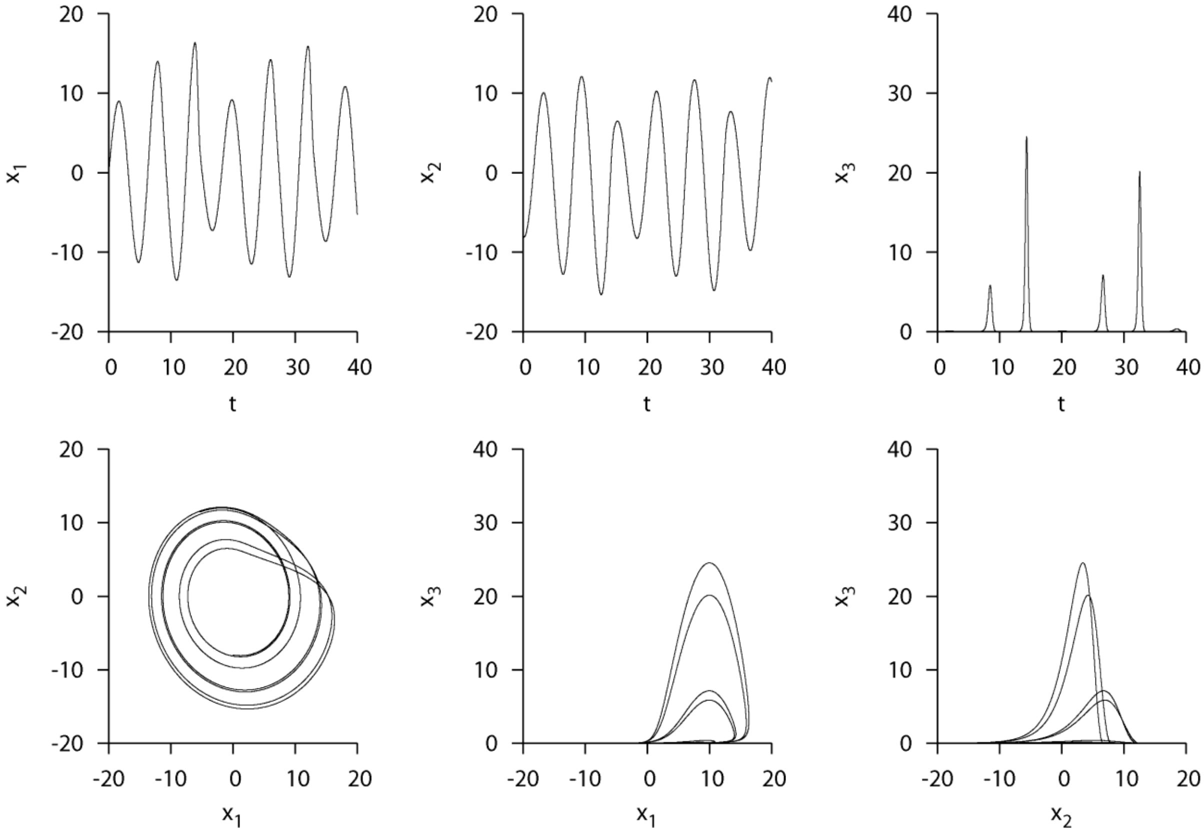

Fig. 1. **Time series and phase portraits of system (20)**

Table 2

Coefficients values for SS (19)

Coefficients	Values					
	Without filtering		$K_S = 6$		$K_S = 3$	
	$\bar{X}$	$\sigma(X)$	$\bar{X}$	$\sigma(X)$	$\bar{X}$	$\sigma(X)$
N_0	0.03000	0.00022	0.03000	2.53569·10-5	0.03000	5.83039·10-6
N_1	-10.00284	0.13306	-10.00373	0.00665	-10.00377	0.00051
N_2	0.48028	0.01973	0.48041	0.00099	0.48042	7.58047·10-5
N_3	-9.84989	0.13291	-9.85079	0.00664	-9.85083	0.00051
N_4	1.98472	0.07859	1.98543	0.00135	1.98539	0.0001
N_5	-0.29764	0.03158	-0.29786	0.00094	-0.29786	9.12834·10-5
N_6	1.96955	0.14319	1.9707	0.00193	1.97063	0.00011

Table continuation 2

N_7	0.01476	0.003	0.01479	$9.02106\cdot10^{-5}$	0.01479	$7.18708\cdot10^{-6}$
N_8	-0.09838	0.029412	-0.09857	0.00084	-0.09858	$8.58792\cdot10^{-5}$
N_9	-0.0004	0.06561	$4.41435\cdot10^{-5}$	0.00081	$1.82548\cdot10^{-5}$	$2.26824\cdot10^{-5}$
N_{10}	-0.09847	0.00542	-0.09854	0.0001	-0.09853	$8.43562\cdot10^{-6}$
N_{11}	0.01477	0.00333	0.01478	0.00013	0.01478	$1.1944\cdot10^{-5}$
N_{12}	-0.09844	0.01019	-0.09854	0.00017	-0.09853	$9.42224\cdot10^{-6}$
N_{13}	$4.5247\cdot10^{-7}$	0.00042	$-4.61089\cdot10^{-7}$	$1.82978\cdot10^{-5}$	$-6.35548\cdot10^{-7}$	$1.85519\cdot10^{-6}$
N_{14}	$-8.37181\cdot10^{-6}$	0.00325	$3.76201\cdot10^{-6}$	0.00013	$4.53656\cdot10^{-6}$	$1.25883\cdot10^{-5}$
N_{15}	$3.57872\cdot10^{-5}$	0.00483	$-4.67331\cdot10^{-6}$	$8.45372\cdot10^{-5}$	$-2.11216\cdot10^{-6}$	$2.81773\cdot10^{-6}$
N_{16}	$-1.59664\cdot10^{-9}$	$1.69664\cdot10^{-7}$	$-4.10451\cdot10^{-11}$	$1.00304\cdot10^{-8}$	$3.28974\cdot10^{-10}$	$2.76239\cdot10^{-9}$
D_0	1	0	1	0	1	0
D_1	-0.09851	0.00135	-0.09852	0.0001	-0.09852	$1.14972\cdot10^{-5}$
D_2	$8.44542\cdot10^{-6}$	0.00191	$-6.44653\cdot10^{-7}$	$9.54622\cdot10^{-5}$	$1.4824\cdot10^{-7}$	$9.84219\cdot10^{-6}$

To determine which of the coefficients are present in the SS, the "significance" value was calculated [11]

$$S(X)=\frac{\bar{X}}{\sigma(X)}. \quad (21)$$

The values of significance for different coefficients of the SS numerator are shown in Fig. 2. It can be seen that for the coefficients N_9, N_{13}, N_{14}, N_{16}, N_{15} the significance index is several orders of magnitude less than for the other coefficients, and the indicated coefficients should be considered equal to zero, and their non-zero values in Table 2 are explained by computational errors. For comparison, the coefficient $N_0=0.03$ cannot be considered equal to zero, since its significance index is on the order of 10^2-10^3, which is much higher than that of the above mentioned zero coefficients. Fig. 3 shows the dependence of the absolute value of the coefficient obtained by solving the system of algebraic equations on the number of the point k from which the formation of the system of equations begins. Graphs are plotted for coefficients N_4 and N_9 in a logarithmic scale. Figure 3 shows that the values of N_9 coefficient obtained for different k may differ by several orders. As a result the standard deviation for a given SF coefficient increases, and the significance indicator decreases. At the same time, the values of the N_4 coefficient found for different k practically do not differ.

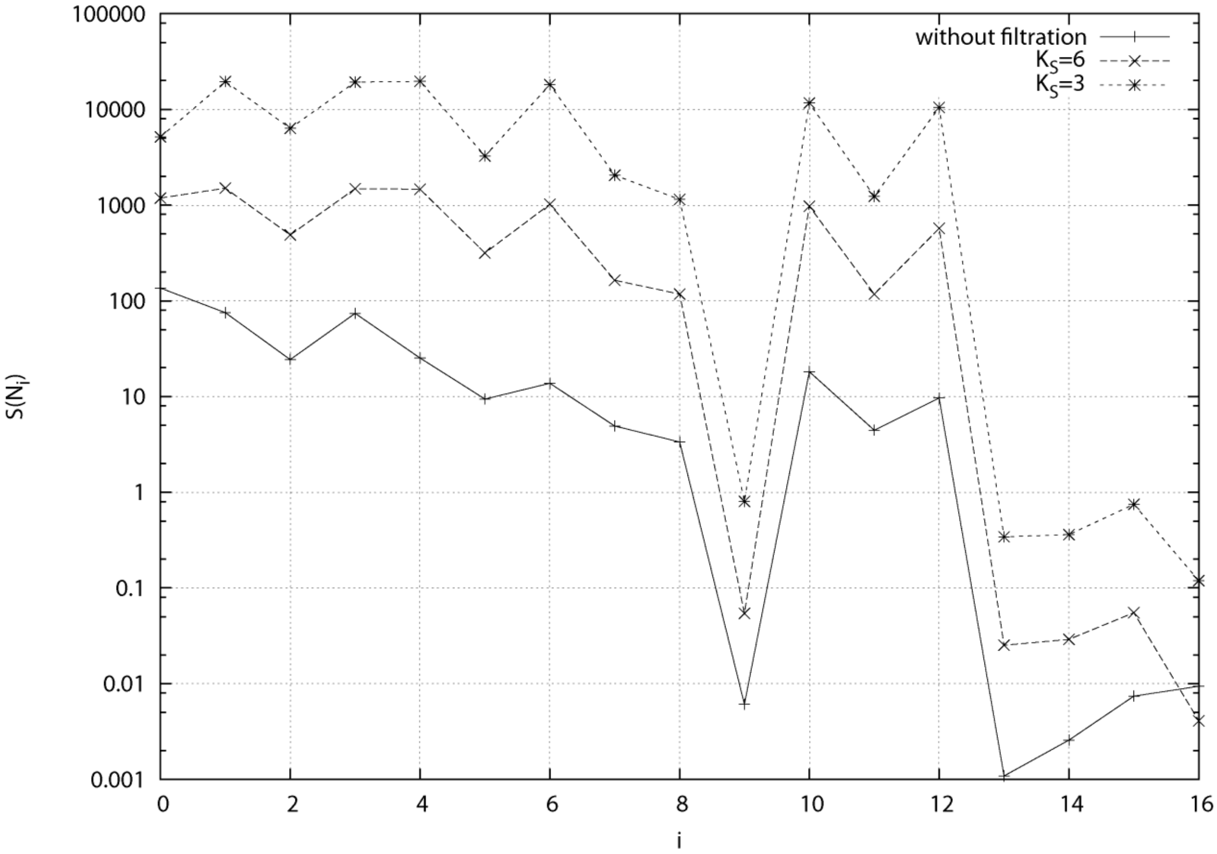

Fig. 2. **Significance index (21) for coefficients of SS (19)**

The effect of Δn value on the results of the algorithm was also investigated. Fig. 4 shows the dependences of the reconstruction quality index $K(\Delta n)$ on a logarithmic scale for the case of the absence of filtering coefficients and filtering with a parameter $K_S = 6$. Fig. 5 shows the dependence of the values of the SS coefficients N_1, N_4, N_9 on Δn. It can be seen that at small Δn values the quality index (Fig. 4) reaches 10^5, which indicates a divergence of numerical integration.

The reason for this is that for small Δn (less than 200 points for the case of Fig. 4), the points at which linear equations are formed are located close to each other, and the corresponding system of equations reflects the nature of not the entire time sequence, but only a small part of it. As a result, the obtained values of the coefficients differ significantly from the real ones, which can be seen in Fig. 4 at $\Delta n < 200$. As Δn increases, the SS coefficients tend do some constant values, and the quality index decreases, indicating that the correct values are found.

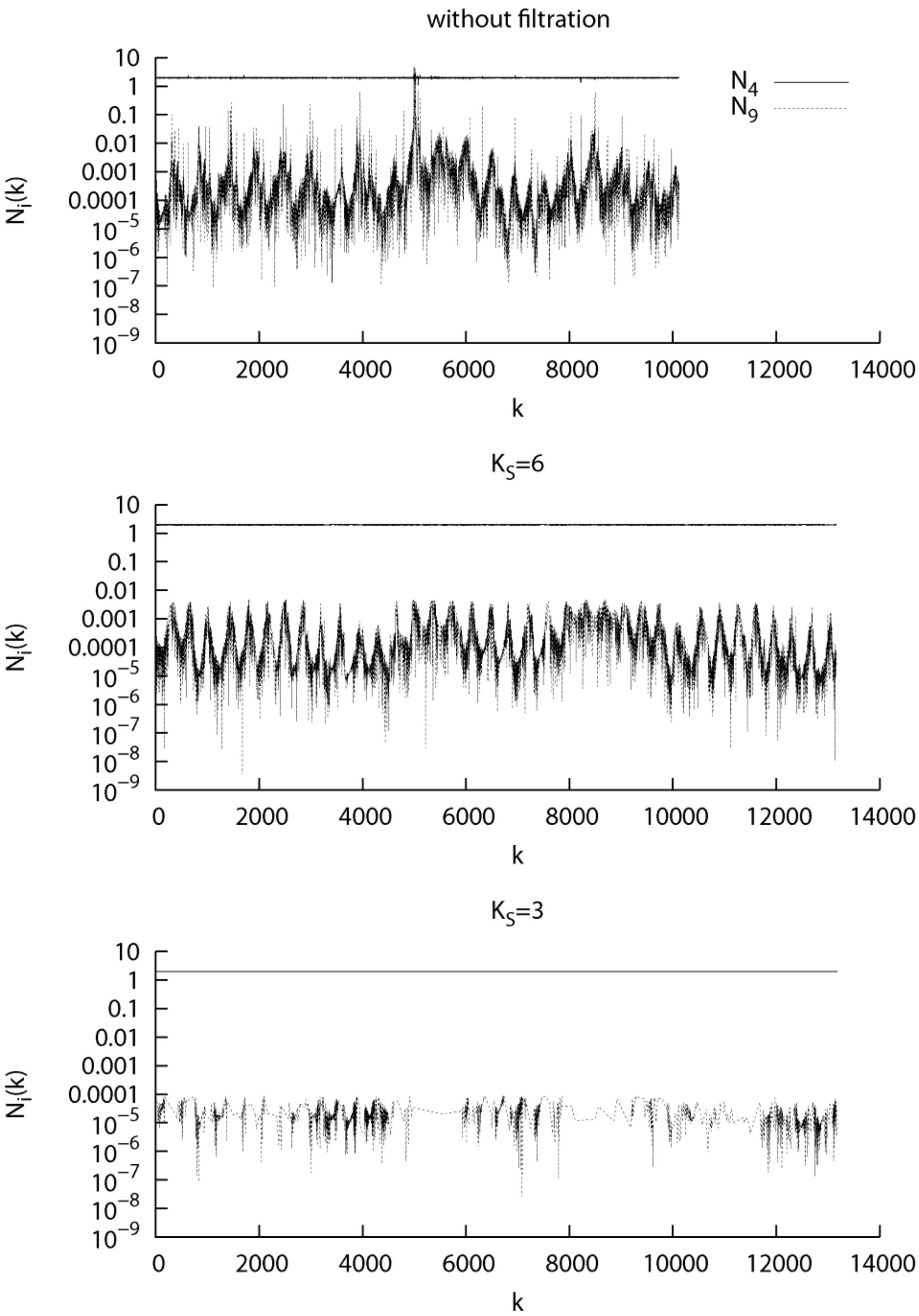

Fig. 3. **Values of SS (19) coefficients depending on the position of the starting point** k

At some points (i.e., at some k) the formed systems of linear equations can have solutions that exceed by several orders of magnitude the solutions at other points (Fig. 3). If used for averaging the solutions of all systems, then this introduces an error in the determination of the SS coefficients, which manifests itself in the form of peaks for some Δn values in the graphs in Fig. 4 and Fig. 5. If the solutions of systems of linear equations are filtered in accordance with condition (12), then the reconstruction quality index decreases (Fig. 4), and the values of the SS coefficients less depend on Δn (Fig. 5).

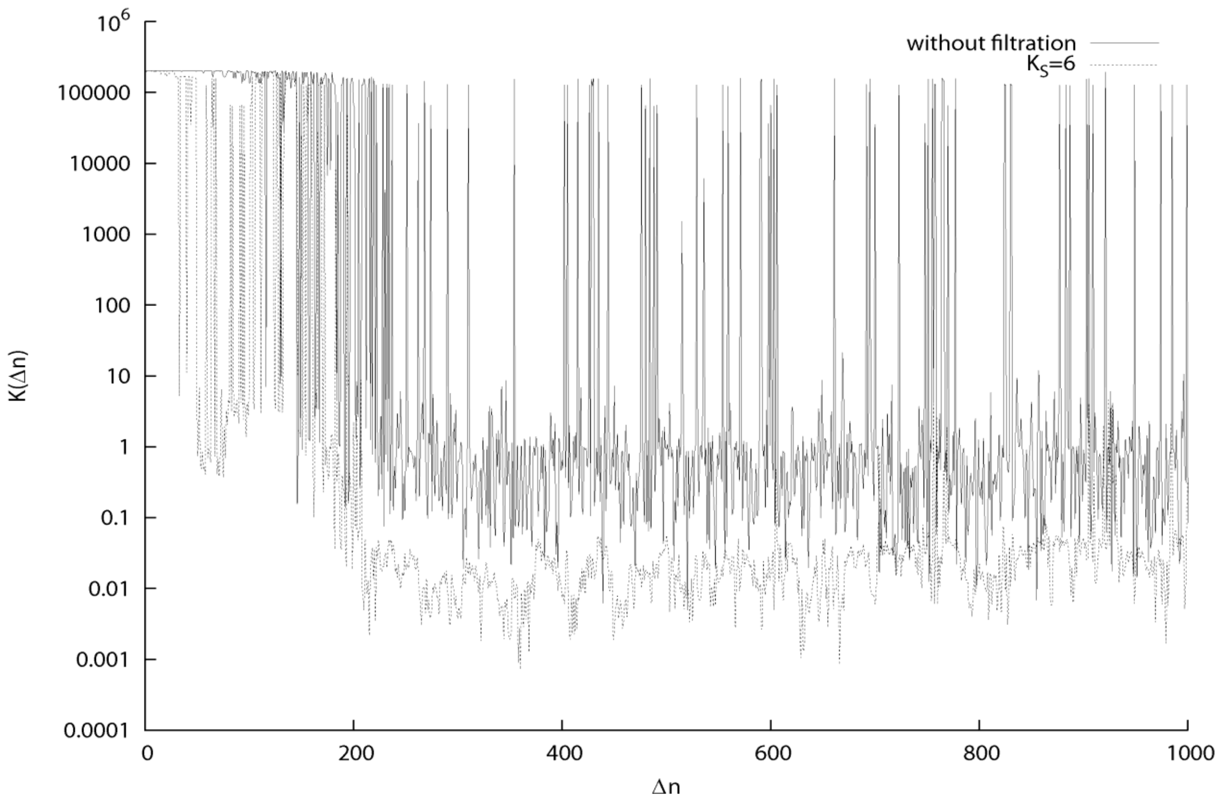

Fig. 4. **Quality index (17) depending on the interval for the formation of the algebraic system** Δn

Conclusions

The article proposed an algorithm for calculating the SS coefficients based on the time series of one observed OS variable. Due to the filtration of systems of linear equations solutions, the precision of calculating the coefficients is increased. The possibility of choosing the most adequate solution from the set of possible ones is provided by varying the sampling step of points for compiling systems of linear equations and then using the reconstruction quality index to evaluate the solutions obtained. The results presented in the article confirm the increase in the calculations precision when using these improvements. Also, the algorithm can be generalized to the case of a fractional-rational SF with a large number of terms. The article did not consider the effect of noise in the original data on the results, since it is assumed that before applying the algorithm, the noise is removed using one of the known filtering methods. It is also assumed that a noise-resistant differentiation method can be used to eliminate the effect of noise on the process of numerical differentiation.

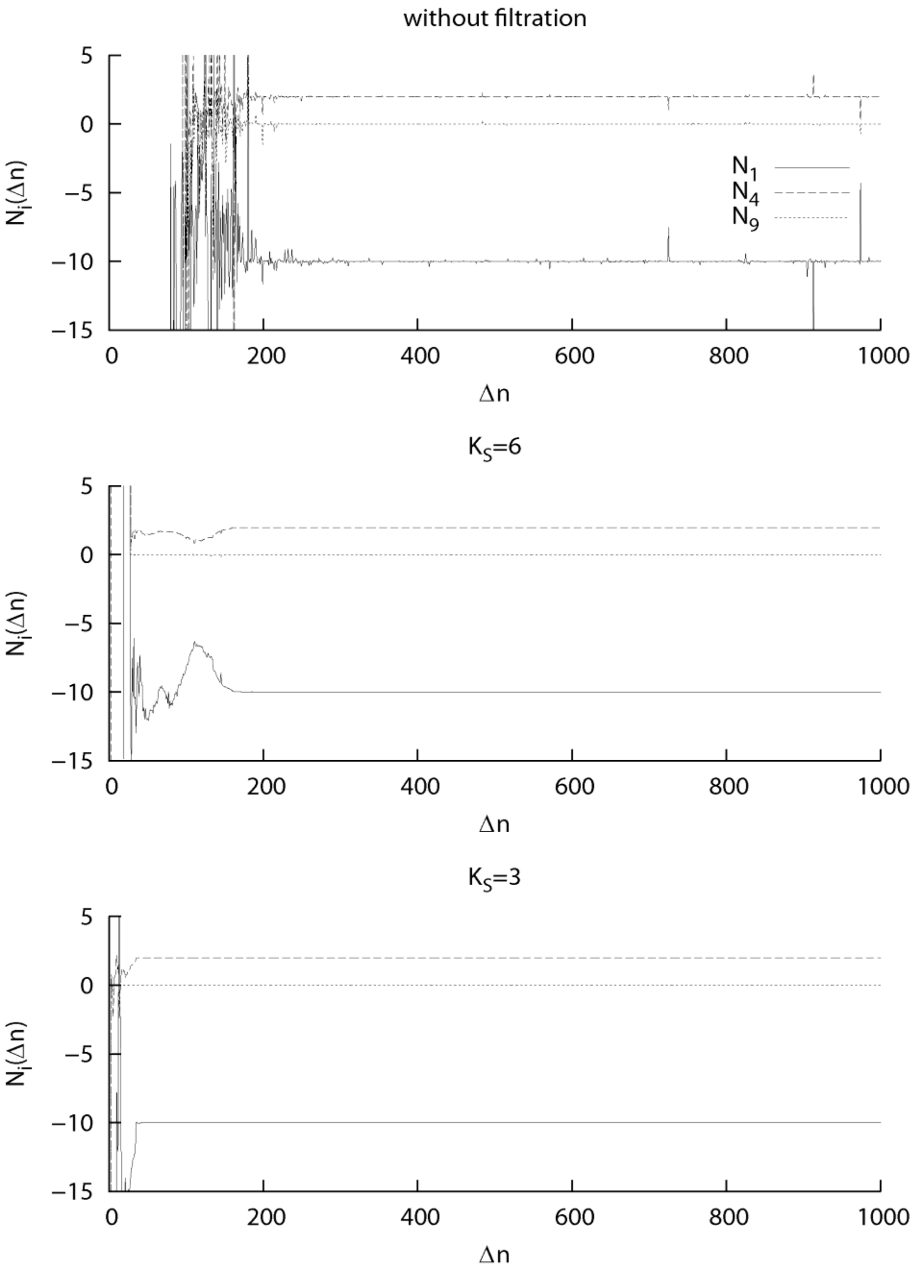

Fig. 5. **Values of SS (19) coefficients depending on the interval for the formation of the algebraic system** Δn

References:

1. J. Cremers, A. Hubler, Construction of differential equations from experimental data, Naturforsch. 42a (1987) 797-802.
2. J. L. Breeden, A. Hubler, Reconstructing equations of motion from experimental data with unobserved variables, Phys. Rev. A 42 (1990) 5817-5826.
3. G. Gouesbet, Reconstruction of standard and inverse vector fields equivalent to the Rössler system, Phys. Rev. A 44 (1991) 6264–6280.
4. F. Takens, Detecting strange attractors in turbulence, in: D.A. Rand, L.S. Young (Eds.), Dynamical System and Turbulence, Lecture Notes in Mathematics, New York: Springer, 898 (1981) 366 – 381.
5. G. Gouesbet, Reconstruction of the vector fields of continuous dynamical systems from numerical scalar time series, Phys. Rev. A 43 (1991) 5321–5331.
6. G. Gouesbet, C. Letellier, Global vector-field reconstruction by using a multivariate polynomial L2 approximation on nets, Phys. Rev. E, V. 49 (1994), no.6, pp. 4955-4972.
7. J. Maquet, C. Letellier, L. A. Aguirre, Scalar modeling and analysis of a 3D biochemical reaction model, Journal of Theoretical Biology 228 (2004) 421–430.
8. V. Gorodetskyi, M. Osadchuk, Reconstruction of chaotic systems of a certain class, International Journal of Dynamics and Control, 3 (2015) N 4, 341-353.
9. L. Le Sceller, C. Letellier, G. Gouesbet, Structure selection for Global vector field reconstruction by using the identification of fixed points, Phys. Rev. E, V. 60 (1999), no.2, pp. 1600-1606.
10. O. E. Rössler, An equation for continuous chaos, Phys. Lett. A 57 (1976) 397–398.
11. C. Lainscsek, C. Letellier, I. Gorodnitsky, Global modeling of the Rössler system from the z-variable, Phys. Lett. A 314 (2003) 409-427.