УДК 681.3.06

N.L. Dorosh, L.N. Petrechuk, T.M. Fenenko

# THE STUDY OF THE TIME SERIES OF SPECIALIZED INFORMATION

***Annotation***. *Excel add-in (XLA) has been developed. Add-in can be used for spectral analysis of time series, data visualization, exponential smoothing, non-linear Box-Cox transformation and smoothing the data using digital filtering algorithms.*

Keywords: *Excel add-in, time series, spectral analysis, impulsive description, cyclic changes.*

## Formulation of the problem

The prevalence and convenience of the Ms Excel spreadsheet has made it one of the most popular software products for performing versatile calculations. Another important advantage is possibility of various programmatic modules to connect to MS Excel as add-in [1]. The Excel add-in (xla), which will allow to execute graphic presentation of data and signal spectral density, also the output of data processing results by the method of digital filtering on the worksheet - compensates for the lack of functions for performing spectral analysis in a table processor.

## Formulating the purpose of the article

The purpose of this work was to create an add-in, modules of which allow performing of the time series smoothing, its transformation by the Box-Cox method and smoothing by digital filter with finite impulse response (FIR) and time series analysis.

The object of the research is time series, which is processed in the projected add-in.

The subject of the research is methods of data smoothing, which are used in the developed add-in.

## Main part

The new add-in provides additional functionality for processing of time series in the MS Excel New add-in provides additional functional possibilities of treatment of sentinel rows in the processor of MS Excel [2] and uses functions that are in the library dll-file. The following features are implemented in the functions: viewing of the time series data; non-linear Box-Cox transformation (Box - Cox Transformation);, smoothing by a digital filter with finite impulse response (FIR).

Functions are implemented in C ++ language in C ++ Builder environment. Standard dll files are created by C ++ Builder compilers [3] and and use the internal programming language VBA (Visual Basic for Applications) in the project. Functions of the standard dll can not be induced directly from the worksheet, or from MS Excel macros sheet [1]. Special interface functions are built for this purpose/

The diagram of components, which leads to the structure of building on of Excel, is given on Figure 1.

The diagram of components, which proves structure of Excel add-in, is shown on Figure 1.

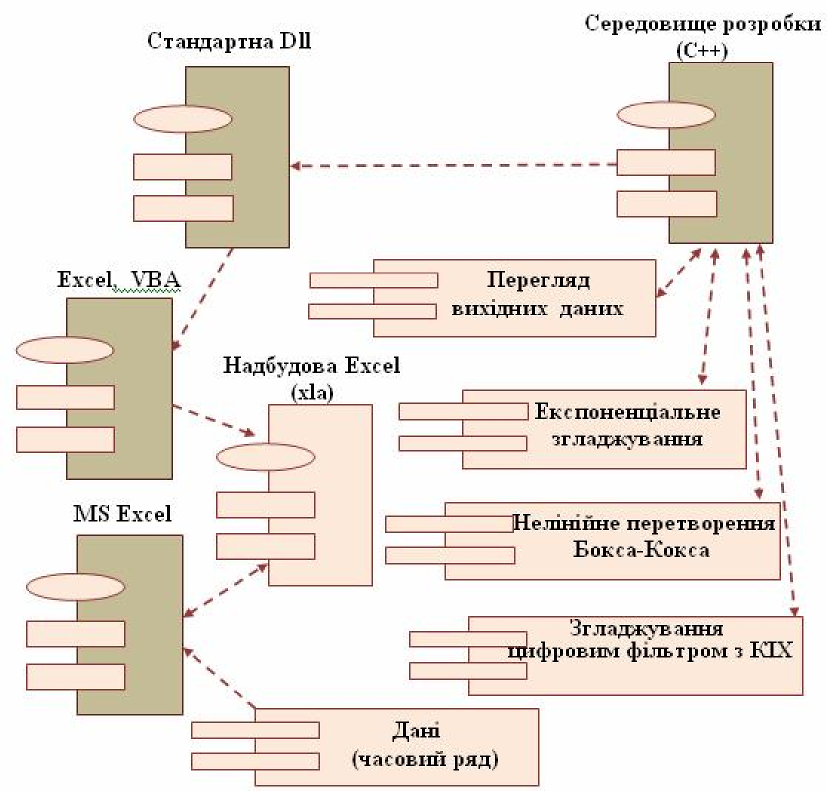



Figure 1. - Diagram of components for Excel add-in

Exponential smoothing of the signal in the module of add-in is carried out according to the recurrence formula:

$$S(t) = \alpha \cdot x(t)+(1 - \alpha ) \cdot S(t-1) \quad (1)$$

where α - is a smoothing parameter; S(t) - smoothed time series value

X = { x1, x2, x3 ..xt}.

The value of the smoothing parameter α is within $0 \le \alpha \le 1$, It is recommended to select the value α from 0.35 to 1.

Nonlinear transformation of signal by the Box-Cox method

Smoothing of time series X={ x1, x2, x3 .. xT} is carried out under a formula:

$$S(t)=\begin{cases}\dfrac{X(t)^{\lambda}-1}{\lambda} & \rightarrow \quad \lambda \neq 0\\ Ln(X(t)) & \rightarrow \quad \lambda = 0\end{cases} \tag{2}$$

where S(t) - smoothed value, λ - parameter.

Maximization of the likelihood function logarithm (maximum log-likelihood function) is one of the methods for determining the optimum value of the lambda parameter:

$$f(x,\lambda)=-\frac{N}{2}\ln\left[\sum_{i=0}^{N-1}\frac{(x_i(\lambda)-\overline{x}(\lambda))^2}{N}\right]+(\lambda-1)\sum_{i=0}^{N-1}\ln(x_i) \tag{3}$$

where N - number of observations ;

x - time series observation;

$\overline{\tilde{o}}$ - average value time series observations

$$\overline{x}(\lambda)=\frac{1}{N}\sum_{i=0}^{N-1}x_i(\lambda) \tag{4}$$

That is, it is necessary to choose such lambda parameter value, at which this function takes a maximum value.

In the developed smoothing add-on by linearly filtering of output data with digital filter with finite impulse response is implemented in several ways: averaging over n points (n=3-15); method of parabola through a few equidistant values of signal, with the further use in quality of the smoothed out size of value of parabola in a middle point (n= 5, 7, 9, 11, 13); av of Spenser for 15 and 21 points [4].

Smoothing of the signal by three points was carried out according to the formula:

Y(n)= X(n - 2) + X(n - 1) + X(n) + X(n+1) + X(n+2))/3 (5)

Smoothing by the Spencer method on the 21st point was carried out according to the formula:

$$x_i=\frac{1}{320}(-3x_{i-7}-6x_{i-6}-5x_{i-5}+3x_{i-4}+21x_{i-3}+46x_{i-2}+67x_{i-1}+74x_i$$
$$+67x_{i+1}+46x_{i+2}+21x_{i+3}+3x_{i+4}-5x_{i+5}-6x_{i+6}-3x_{i+7}) \tag{6}$$

Spectral analysis of the signal is intended for the estimation of frequency composition of discrete signal [5]. Nonparametric methods are based on the calculation of estimates of the spectral power density (SPD) immediately after counting the initial sequence, determines their main advantages - possibility of applying to a wide class of stationary signals and noise and high computational efficiency due to the application of Fast Fourier Transform (FFT) algorithms.

For the estimation of he spectral power density, the periodogram method, is used [6], which main point is in the calculation of estimation of SPD of eventual casual sequence of length of N, wis called the periodogram.

$$S(\omega) = \frac{\left|X(e^{j\omega T})\right|^2}{Nf_Д} \tag{7}$$

where $X(e^{j\omega T})$ - spectral density of a finite sequence x (n).

$$X(e^{j\omega T}) = \sum_{n=0}^{N-1} x(n)e^{-j\omega Tn} \tag{8}$$

For diminishing of spectrum spreading effect at the calculation of periodogram by means of discrete Fourier transform (DFT) and, as a result, smoothing of periodogram, weigh functions (windows) are applied, and the modified periodogram, called the modified periodogram, becomes:

$$\hat{S}_w(\omega) = \frac{\frac{1}{f_Ä}\left|X_w(e^{j\omega T})\right|^2}{\sum_{n=0}^{N-1}\left|w(n)\right|^2} \tag{9}$$

where w(n) - weight function (window) of length N;

$X_w(e^{j\omega T})$ - spectral density of multiplication of x(n)·w(n) :

$$X_w(e^{j\omega T}) = \sum_{n=0}^{N-1} x(n)w(n)e^{-j\omega Tn} \tag{10}$$

The window function of the FFT is used to suppress edge discontinuity effects of the discontinuity of real functions by introducing weighting coefficients for fetching data in the window, which provide a reduction in the amplitude of the edge points (start and stop) and, as a result, improve the results of the FFT. Different types of window functions (rectangular window, Hemminga, Hanninga, Blackmana-Harrisa) give different results, both in accuracy, and in frequent resolution and apply for various kinds of signals [7].

The Hamming window is applied in this work.

Let X={of x1, x2, x3 .. xN} - time series which is given, Y={of y1, y2, y3 .. yN} - - time series (signal) after smoothing.

Convolution of the signal has the form:

$$y(n) = \sum_{m=0}^{\infty} h(n-m)x(m) = \sum_{m=0}^{\infty} h(m)x(n-m) \tag{11}$$

where m - delay of sequence, h(n) - impulsive description (IX).

H(z) - z -Figure of IX  of h(n) - calculated by a formula:

$$H(z) = \sum_{n=0}^{\infty} h(n) z^{-n} \tag{12}$$

The transmission function of H(z) is :

$$H(z) = \frac{Y(z)}{X(z)} \tag{13}$$

where X(z), Y(z), - z  - The image of the signal before and after processing.

Let $H\left(e^{\hat{j}\hat{\omega}}\right)$ - Fourier - an image of IX of h(n), where is normalized frequency.

For the frequency description and the transfer function H (x), the following relation is true:

$$H\left(e^{\hat{j}\hat{\omega}}\right) = H(z)\Big|_{z=e^{\hat{j}\hat{\omega}}} \tag{14}$$

In the created add-in, an analysis of the time series, reflecting the number of citizens, who for the first time appealed to the treatment and prophylactic institutions of Ukraine with complaints about the digestive system, is carried out. Output data - monthly amount of patients, who applied for the period from January 2006 to December 2015.

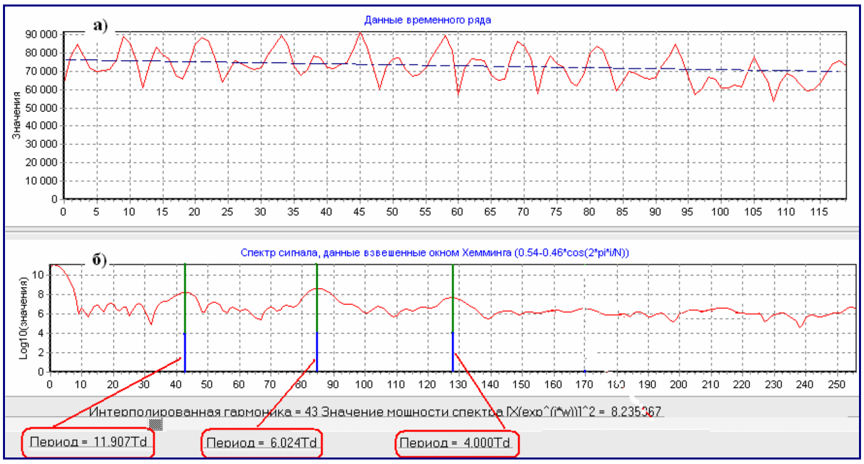



Figure 2. - Time series showing the figures for the 10 years.

The graphic image of the time series under study is shown on Figure 2 and gives the the initial idea of the changing statistics (number of patients) over the past 10 years. Visually on the chart, you can select:

- The long-term trend - The dashed line goes gradually downwards (Figure  2a);

- Cyclic changes which are usually related to the vibrations of physiology activity of of the human body (Figure 2a);

- Seasonal fluctuations (Figure of 2b);

- Irregular fluctuations associated with unpredictable random events (Figure 2b).

On the Figure 2-b the signal spectrum (smoothed by the Hamming window) is represented, harmonics, which correspond to the values of the periods of 12, 6 and 4. Based on the fact that the time series, which is analyzed, consists of monthly indications Coming from that a sentinel row which is analysed consists of monthly testimonies (I.e., has a discretization step in time Td = 1), ehe figures indicate that the arrival of citizens in medical institutions has a certain pattern, has three different time periods: : annual (12 months) semi-annual (6 months) and seasonal (winter, spring, summer, autumn - 4 months).

Graphical representation of data series after processing by the method of exponential smoothing (in the add-in) for different values of the parameter □ is given in the Figure 3.

The results of smoothing of data series in the form of an Excel spreadsheet and charts of residues are presented on Picrure 4.

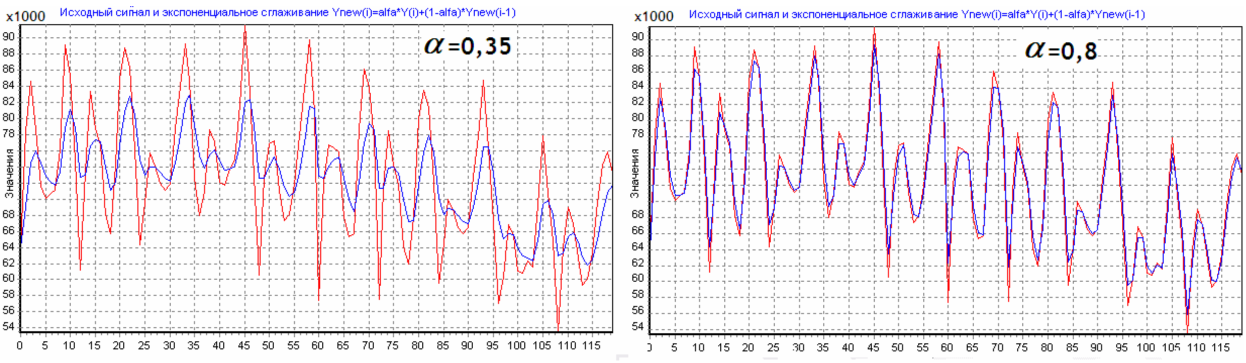
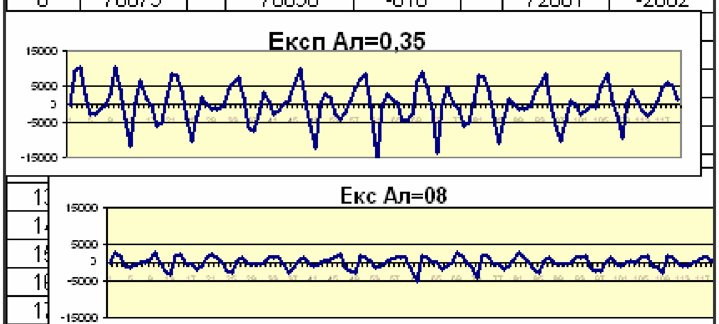



Figure 3. - Method of exponential smoothing

| Вихідні дані | | | Експоненціальне згладжування сигналу | | | |
|---|---|---|---|---|---|---|
| № з/п | Кількість | | Ал=0,8 | Залишки | Ал=0,35 | Залишки |
| 1 | 63985 | | 63985 | 0 | 63985 | 0 |
| 2 | 78356 | | 75482 | 2874 | 69015 | 9341 |
| 3 | 84552 | | 82738 | 1814 | 74453 | 10099 |
| 4 | 78560 | | 79396 | -836 | 75891 | 2670 |
| 5 | 71603 | | 73161 | -1559 | 74390 | -2787 |
| 6 | 70079 | | 70696 | -616 | 72881 | -2802 |




| | | | | | |
|---|---|---|---|---|---|
| 118 | 74198 | 73001 | 1196 | 68158 | 6039 |
| 119 | 75836 | 75269 | 567 | 70845 | 4991 |
| 120 | 73240 | 73646 | -406 | 71684 | 1556 |

Figure 4. - The results of exponential smoothing on Excel sheet

 ISSN 1562-9945

Authenticity of any results is done by comparison. Similar processing of data in relation to the appeal of to medical institutions was carried out in the application ITSM2000 application [8].

Figure 5 shows the screen types of ITSM2000 applications and add-ins for Excel algorithm of nonlinear Box-Cox transformation. The results are alike in the graphical form. Comparative analysis of the Figures is qualitative.

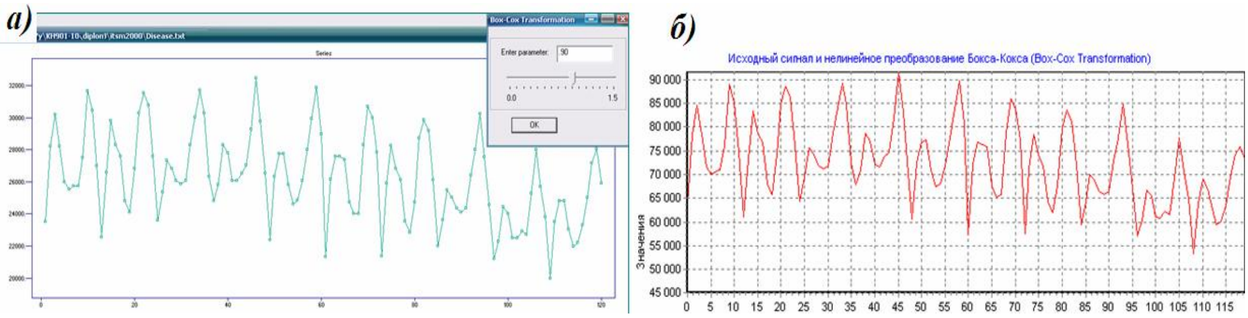



Figure 5. - Non-linear Box-Cox transformation, а) - in ITSM 2000, б) - in add-in

Figure 6 shows the periodogram that was made in ITSM 2000 program. Peak at the area of a certain frequency ω0 indicates that corresponding harmonious component is present in in the spectral decomposition of the autocorrelation function. The higher and sharper the peak is allocated, the bigger part of the power is concentrated near the frequency ω0 and the more important role this frequency plays in the description of the corresponding random process [9].

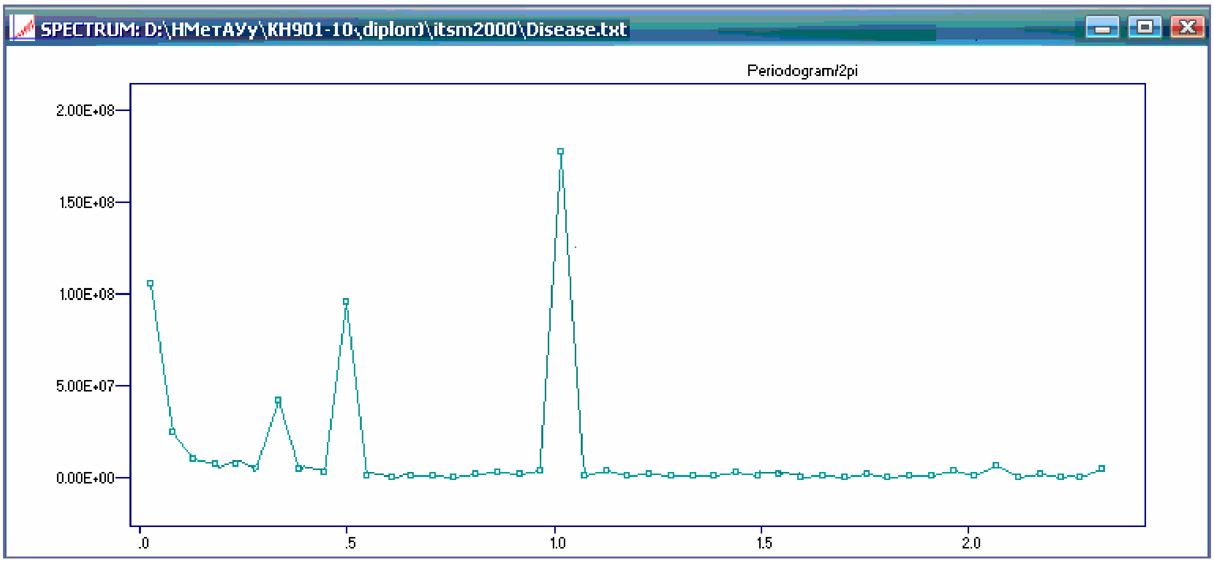



Figure 6. - Periodogram in ITSM 2000

When considering time series per unit of time, one month is taken (one observation per month). On periodogram (Figure 6) on the X-axis argument corresponds to the following values: 0.25 - 3 months, 0.5 - 6 months, 1.0 - 12 months.

The presence of peaks in the range of ≈0.33, 0.5, and 1.0 indicates that there are three cycles in the investigated series:

- - 4 months (0.33 • 12 months = 4 months);
- half-year cycle (0,5 · 12 months = 6 months);
- annual cycle (1,0 · 12 months = 12 months).

The signal spectrum, obtained after processing the time series in the add-in (Figure 7), Shows the presence of three clearly expressed cycles with periods of 4, 6 and 12 months. So, the spectral analysis in the add-in and in the ITSM 2000 program is identical.


Figure 7. - The signal spectrum in the "add-in"

Quantitative analysis of the comparison of the smoothing results, implemented in the created add-in algorithms, is presented on Figure 8.

| A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|
| № з/п | Сигнал, disease | Надбудова | ITSM 2000 | Δ | Надбудова | ITSM 2000 | Δ |
| | | Експон. згладжування, α=0.8 | | | Перетворення Бокса-Кокса, λ = 0.84 | | |
| 1 | 63985 | 63985,39 | 63985,39 | 0,00 | 12965,48 | 12965,48 | 0,00E+00 |
| 2 | 78356 | 75482,10 | 75482,10 | 0,00 | 15371,25 | 15371,25 | 0,00E+00 |
| 3 | 84552 | 82738,12 | 82738,12 | 0,00 | 16386,03 | 16386,03 | 0,00E+00 |
| 4 | 78560 | 79395,84 | 79395,84 | 0,00 | 15404,86 | 15404,86 | 0,00E+00 |
| 5 | 71603 | 73161,33 | 73161,33 | 0,00 | 14250,34 | 14250,34 | 1,20E-13 |
| 6 | 70079 | 70695,65 | 70695,65 | 0,00 | 13995,19 | 13995,19 | 0,00E+00 |
| 7 | 70645 | 70655,10 | 70655,10 | 0,00 | 14090,04 | 14090,04 | 1,70E-13 |
| 8 | 71039 | 70962,19 | 70962,19 | 0,00 | 14156,03 | 14156,03 | 0,00E+00 |
| 9 | 76173 | 75130,91 | 75130,91 | 0,00 | 15010,65 | 15010,65 | 0,00E+00 |
| 10 | 89123 | 86324,19 | 86324,19 | 0,00 | 17126,95 | 17126,95 | 3,80E-13 |
| 11 | 85314 | 85515,92 | 85515,92 | 0,00 | 16509,95 | 16509,95 | 0,00E+00 |
| 12 | 74650 | 76822,88 | 76822,88 | 0,00 | 14758,05 | 14758,05 | 0,00E+00 |
| 116 | 63324 | 62671,74 | 62671,74 | 0,00 | 12852,86 | 12852,86 | 0,00E+00 |
| 117 | 69601 | 68215,43 | 68215,43 | 0,00 | 13914,97 | 13914,97 | 4,60E-13 |
| 118 | 74198 | 73001,19 | 73001,19 | 0,00 | 14682,94 | 14682,94 | 0,00E+00 |
| 119 | 75836 | 75269,15 | 75269,15 | 0,00 | 14954,85 | 14954,85 | 0,00E+00 |
| 120 | 73240 | 73645,72 | 73645,72 | 0,00 | 14523,56 | 14523,56 | 0,00E+00 |

Figure 8. Results of comparative analysis of the functions of the add-in and the package ITSM 2000


ISSN 1562-9945

## Conclusions

The development of the add-on, which can be used to process time series in MS Excel spreadsheets, has been done. The add-in allows to perform a smoothing of the time series and perform a spectral analysis of the data. Output of results on the MS Excel sheet is provided.

Confirmation of the functionality of add-in, which was, created in Excel, was carried out by comparing the results, obtained with a similar processing of the investigated series in the ITSM 2000 program.

The time series of the number of patients with digestive disorders has been studied. Spectral analysis revealed the presence of cyclicity in the investigated series, that is proved with numbers:

- Period = 4Td indicates a presense of four-month cycle;
- Period= 6Td indicates a presense of six-month cycle;
- Period=12Td indicates that the nature of the curve, which reflects the annual number of citizens, who applied to health facilities, is identical for all ten intervals. The obtained results are confirmed by expert evaluations.

## LITERATURE

1. Гайдышев И.П. Решение научных и инженерных задач средствами Excel, VBA и C/C++ / И.П. Гайдышев – СПб.: БХВ – Петербург, 2004. – 512 с.

2. Дорош Н. Л., Храпач Ю. О. Результати розробки надбудови MS Excel для обробки часових рядів // Системні технології. Регіональний міжвузівський збірник наукових праць. – Випуск 2 (97). – Дніпропетровськ, 2015. – С. 8-12.

3. Архангельский А.Я. Программирование в C++Builder 6 / А.Я. Архангельский – М.: издательство «Бином», 2008. – 1152 с.

4. Отнес Р. Прикладной анализ временных рядов. Основные методы / Р. Отнес, Л. Эноксон. Пер. с англ. под ред. В. И. Хохлова. – М.: Мир, 1982. – 428 с.

5. Дженкинс Г. Спектральный анализ и его приложения. Выпуск 1 / Г. Дженкинс, Д. Ваттс – М.:Мир, 1971. – 462 с.

6. Солонина А.И. Цифровая обработка сигналов и MATLAB: учеб. пособие / А. И. Солонина, Д. М. Клионский, Т. В. Меркучева, С. Н. Перов. – СПб.: БХВ-Петербург, 2013. – 512 с.

7. Лайонс Р. Цифровая обработка сигналов / Р.Лайонс – М.:Бином, 2006. – 656 с.

8. ITSM 2000 Professional Version 6.0, developed by Peter J. Brockwell and Richard A. Davis, B&D Enterprises, Inc., Copyright 1999.

9. Хемминг Р.В. Цифровые фильтры. / Р.В. Хемминг, пер. с англ. под ред. А.М. Трахтмана. – М.: Сов. радио, 1980.- 224 с.

Information is about authors

Dorosh Natalia Leonidovna, associate professor, NMетAU, department of information technologies and systems.

Petrechuk Lina Nikolaevna, senior teacher, NMетAU, department of the applied mathematics and computing engineering.

Fenenko Tatyana Mikhailovna, senior teacher, NMетAU, department of the applied mathematics and computing engineering.

## ABSTRACTS

УДК 681.3.06

Дорош Н. Л., Петречук Л.М., Фененко Т.М. Дослідження часового ряду спеціалізованої інформації // Системні технології. Регіональний міжвузівський збірник наукових праць. – Випуск ? (??). – Дніпро, 2017. – С. ??-??.

Представлені результати дослідження часового ряду спеціалізованої інформації у створеній надбудові MS Excel, яка дозволяє виконати його спектральний аналіз. Надбудова містить модулі перегляду даних у часовій та частотній областях. Проведено порівняльний аналіз отриманих результатів.

Бібл. 9, іл. 8, формул 14.

УДК 681.3.06

Дорош Н. Л., Петречук Л. Н., Фененко Т.М. Исследование временного ряда специализированной информации // System технології. Regional міжвузівський collection of scientific labours. it is Producing ? (??). it is Dnepr, 2017. - С. ??-??.

Представлены of результаты исследования временного ряда специализированной информации in созданной надстройке of MS Excel, которая позволяет выполнить его спектральный анализ. Надстройка of включает модули просмотра данных во временной и частотной areas. Проведен of сравнительный анализ полученных результатов.

Библ. 9, ил. 8, formulas 14.

УДК 681.3.06

Dorosh N. L., Petrechuk L.N., T.M. Fenenko The of study of the time series of specialized information // System технології. Regional міжвузівський collection of scientific labours. it is Producing ? (??). it is Dnepr, 2017. - С. ??-??.

The of results of time series specialized information research are submitted in of MS Excel add - in that allows you to perform its spectral analysis. Add - ins include viewing data in the time and frequency domains. An of comparative analysis of the results has been conducted.

Bibl. 9, il. 8, formulas 14.