Method of neural networks committees in calculation of time series maximal Lyapunov exponents

L. A. Dmitrieva, S. S. Chepilko, Yu. A. Kuperin
Department of Physics, Saint-Petersburg State University, Russia;
e-mail: ludmila.dmitrieva@gmail.com

The present paper deals with modification of the neural networks estimation method of the maximal Lyapunov exponent (MLE) for chaotic time series [1] and development of this method for processing series with real world application. Namely, the necessity to use committees of neural networks for MLE calculation is strongly grounded. The method of estimating the Lyapunov exponent calculation error is elaborated. The technique of forecasted trajectories divergences averaging on the delayed pseudoattractor of time series is introduced. The separation of two important cases when MLE is zero and the case when it is small but positive is accomplished by making use of appropriate statistical tests. It is shown that even the modified method of neural networks committee MLE estimation can give positivity of MLE for stochastic series at relatively high statistical characteristics of the neural networks forecasts quality. Additional tests and researches for identification chaos in time series are required. The proposed approach is tested on the model chaotic and periodic time series as well on time series having real world application such as EEG signals and tensotremorogram signals.

1 Introduction

In the present paper we propose the modification of the method of neural network estimation of the maximal Lyapunov exponent (MLE) for the chaotic time series [1] and develop it for processing series of an arbitrary nature. The necessity of serious modification and development of the above method is motivated by the following checked fact. The positivity property of the maximal Lyapunov exponent calculated by the neural network approach proposed in [1], combined with the high statistical characteristics of the neural network forecasts quality, can not be considered as identification of the chaos in the considered time series. We exemplify below that for stochastic time series MLE calculated by the modified neural network approach is positive in the range of MLE miscalculation. At the same time statistical characteristics of the neural network forecasts quality are relatively high. This implies that identification of chaos requires additional test and researches.

Another motivation of the present paper is the following. Being productive in general the existing form [1] of the neural network approach for calculating the maximal Lyapunov exponent suffers from serious shortcomings preventing it to become reliable method of MLE calculating directly from time series. In the present paper to overcome the weak points we propose to use committees of neural networks for providing stability of the MLE calculation. The use of committees also allows to estimate the Lyapunov exponent calculation error. We also propose the technique of averaging of forecasted trajectories divergences on the delayed pseudoattractor of time series. The paper also cosideres two important cases when the MLE is supposed to be zero and the case when it is small but positive. Separation of these two cases is realized by making use of appropriate statistical tests.

To our strong belief further development of the method will allow to provide possibility of accurate identification of the time series under consideration. The latter problem without any doubts is extremely important for practicians dealing with time series with real world applications, such as EEG signals, tensotremorograms, financial time series.

2 Notion of Maximal Lyapunov Exponent

Below we briefly explain a concept of the maximal Lyapunov exponent on "physical level of rigor". Let us pick out two arbitrary points lying on the dynamical system attractor $x(t_0)$ and $\tilde{x}(t_0)$ These points correspond to the time t_0, and the distance between points is $\|x(t_0) - \tilde{x}(t_0)\| = \epsilon(t_0)$, where $\epsilon(t_0)$ is some positive number. In some time interval Δt these points evolve to the points $x(t)$ and $\tilde{x}(t)$, where $t = t_0 + \delta t$, and the distance between them will be $\|x(t) - \tilde{x}(t)\| = \epsilon(t)$. Let us assume that points diverge exponentially in time

$$\epsilon(t) \cong \epsilon(t_0) e^{\lambda \Delta t} \tag{1}$$

where λ is some parameter which we shall call as maximal Lyapunov exponent.

The latter equation gives λ as

$$\lambda \cong \frac{1}{\Delta t} \ln \frac{\epsilon(t)}{\epsilon(t_0)} \tag{2}$$

However to determine λ exactly one have to turn to the limit $\Delta t \to \infty$ and after that consider the average over the points $x(t_0)$ belonging to the dynamical system attractor. One also have to require that at $\Delta t \to \infty$ the quantity $\epsilon(t)$ has to be not greater than the size of the attractor. It gives the following definition of maximal Lyapunov exponent λ:

$$\lambda = E\left[\lim_{\Delta t \to \infty,\ \epsilon(t) < diam(A)} \frac{1}{\Delta t} \ln \frac{\epsilon(t)}{\epsilon(t_0)}\right] \tag{3}$$

Here L is the attractor of dynamical system, $diam(A)$ is the size of attractor, and $E[.]$ is the average over the distribution of points $x(t_0) \in A$.

So the maximal Lyapunov exponent describes the exponential increasing of the distance between time series trajectories. In this content the positive maximal Lyapunov exponent is the indicator of chaotic regime in the system dynamics.

Among the numerical algorithms known for calculating the maximal Lyapunov exponent directly from time series let us mention the Wolf [2] algorithm based on the Benettin method [3] as well as Sano& Sawada [4] and Eckmann [5] algorithms. Separately stand algorithms by Rosenstein, Collins& Luca [6] and Kantz [9], [7].

3 Algorithm of Neural Network Estimating the Time Series Maximal Lyapunov Exponent Proposed in [1]

In this section we briefly describe the method of neural network estimation of the maximal Lyapunov exponent of the chaotic time series proposed in [1] and expose its imperfect features which demand serious modification and development of the approach.

The first step of neural network estimation of MLE is training the neural network on the time series under consideration. In [1] as time series was taken the dependence of one phase coordinate of the known chaotic dynamical system on the time variable. In what follows such series will be called the dynamical ones. As inputs for neural network traditionally were used several lags of time series. In the case of dynamical time series according to the Takens theorem [8] the attractor reconstructed in the lag space by the time delay method provided the embedding dimension is enough preserves the most important dynamical and geometrical properties of the initial dynamical system, in particular, it preserves the Lyapunov exponents. The neural network solves the problem of approximating the mapping F in the delay space:

$$x(t+i\tau) = F(x(t+(i-1)\tau), x(t+(i-2)\tau), ...$$

$$..., x(t+(i-m)\tau)),\ i = 1, ..., n \tag{4}$$

where m - is the embedding dimension, τ - time delay and n - is the amount of training sampling. Different methods of estimation of embedding dimension and time delay are described, for instance, in [7].

Provided the neural network is trained with sufficiently high quality one chooses from time series arbitrary point $x(t_i)$ and forms the lag vector of time delayed values $(x(t_i), x(t_i-\tau), ..., x(t_i-(m-1)\tau))$. Using the components of this lag vector as inputs for neural network one makes forecast and predict the value $x(t_i+\tau)$ of the time series. This forecast we shall denote as $x^{(f)}(t_i+\tau)$. Then one realize multistep forecasting $x^{(f)}(t_i+n\tau)$ on the time horizon H $(n = 1, ..., H)$ as follows:

$$x^{(f)}(t_i+n\tau) = F(x^{(f)}(t_i+(n-1)\tau), ..., x^{(f)}(t_i+\tau),$$

$$x(t_i), ..., x(t_i-(m-n)\tau)),\ n < m, \tag{5}$$

$$x^{(f)}(t_i+n\tau) = F(x^{(f)}(t_i+(n-1)\tau), ...,$$

$$x^{(f)}(t_i+\tau), x(t_i)),\ n = m, \tag{6}$$

and

$$x^{(f)}(t_i+n\tau) = F(x^{(f)}(t_i+(n-1)\tau), ...,$$

$$x^{(f)}(t_i+(n-m)\tau)),\ n > m. \tag{7}$$

Then one introduces a small perturbation of the point $x(t_i)$: $\tilde{x}(t_i) = x(t_i) + a_i$, where $a_i \sim 10^{-8}$ and forms the lag vector of the perturbed point $(\tilde{x}(t_i), x(t_i-\tau), ..., x(t_i-(m-1)\tau))$. Input this lag vector to the neural network and find $\tilde{x}^{(f)}(t_i+\tau)$ with further repetition of the multistep forecasting procedure (5), (6), (7). Thus one obtains forecasts of the perturbed point $\tilde{x}^{(f)}(t_i+n\tau),\ n = 1, ..., H$.

Next one calculates the quantities $\ln d_i(n) = \ln|x^{(f)}(t_i+n\tau) - \tilde{x}^{(f)}(t_i+n\tau)|,\ n = 1, ..., H$ and choose points for which $\ln d_i(n) < 0$.

Finally one makes the plot of $\ln d_i(n),\ n = 1,...,H$ versus the number of iteration n and estimates the corresponding linear regression. Slope of the regression line gives the maximal Lyapunov exponent.

In [1] the proposed approach is illustrated by two examples: the Lorentz and Henon systems.

Being productive in general, the mentioned form of the neural network approach for calculating the maximal Lyapunov exponent suffers from following disadvantages.

The results of calculating essentially depends on the choice of the time series point $x(t_i)$.

Each training of the neural network even with the fixed architecture leads to different results in MLE calculating.

Error of the maximal Lyapunov exponent calculation is not discussed at all. Moreover, the most obvious from the first glance making use of the standard error of the regression coefficient estimation is not adequate for the MLE error estimation, since the result depends on a lot of parameters of the method.

From [1] it is not clear how to deal with the case when MLE is in the vicinity of zero. In [10] authors claim that they considered sine to verify that for this time series MLE is zero. However no concrete details and discussions are given. To our opinion the problem of distinction cases when MLE is zero (for periodic and quasiperiodic data) and the case when it is small but positive is of great importance, since a grate number of chaotic systems has very small MLE.

We exemplify below that for stochastic time series MLE calculated by the modified neural network approach is positive in the range of MLE miscalculation. At the same time statistical characteristics of the neural network forecasts quality (such as R^2 and mean-square error (MLE)[11]) are relatively high. This means that positivity property of the maximal Lyapunov exponent calculated by the neural network approach proposed in [1], combined with the high statistical characteristics of the neural network forecasts quality, can not be considered as identification of the chaos in the considered time series. This implies that identification of chaos requires additional test and researches.

4 Modification and Development of Algorithm

In order to obtain the time series forecasts we chose multilayer perceptron as the appropriate neural network. The perceptron has been trained by Levenberg-Markwardt learning algorithm [11]. In order to train the chosen neural network we have splitted the sample into the training set, validation set and test set. The number of the training epochs has been determined by the numerical experiments. To estimate the quality of the trained neural network operation we have used the following statistics: mean square error (MLE) and multiple regression coefficient R^2 [11] calculated on each subset of the sample.

In order to improve the mentioned above statistics and to overcome one of the main shortcomings of the approach [1] described in previous Section we used the committee of neural networks [12]. The trained neural networks have been included into the committee if they pass through the "quality test", i.e. statistics for each neural network is greater than threshold with respect to multiple regression coefficient R^2. In what follows we call this threshold as "NN trust level". We took it equal to 0.8 or 0.9. The neural network with R^2 larger than "NN trust threshold" participated in the further calculation of the maximal Lyapunov Exponent λ_q. Below q is the number of the successful neural network, $q = 1,...N$. Maximal Lyapunov exponent λ of time series under consideration is obtained by averaging all successfully calculated Lyapunov exponents λ_q: $\lambda = \sum_q \lambda_q/N$ over the neural networks included into committee. The MLE estimation error $\Delta\lambda$ was calculated as standard deviation $\Delta\lambda = \sqrt{D(\lambda)}$, where $D(\lambda)$ is variance of the sample $\{\lambda_q\}_{i=1}^N$.

Another serious modification proposed in the present paper is that we average for each forecast iteration n the divergences of perturbed and unperturbed trajectories $s_q^i(n) = x_q^f(t_i+n\tau) - \tilde{x}_q^f(t_i+n\tau)$ with respect to "good" sampling points $x(t_i)$. We call the time series point $x(t_i)$ as "good" if it "generates" regression line with determination coefficient R^2 larger than established "Regression trust threshold". In more details, we construct the linear regression of $\ln(d_q^i(n))$, where $d_q^i(n) = |s_q^i(n)|$, versus the forecast iteration number n, $n = 1,...,H$, where H is the horizon of multistep forecasts. We estimate the MLE λ_q^i "generated" by the time series point $x(t_i)$ as the slope of regression line (index q means the number of the used neural network in committee). Then we calculate the determination coefficient

$$R^2 = \frac{D(\hat{y})}{D(y)} = 1 - \frac{D(e)}{D(y)}. \quad (8)$$

Here $y(n) = d_q^i(n) = \ln|s_q^i(n)|$ are dependent regression variables, $\hat{y}(n) = a^i + \lambda_q^i n$ are regression values and $e(n) = y(n) - \hat{y}(n)$ are regression errors.

The averaging of the divergences of perturbed and unperturbed trajectories for each forecast iteration n is made with the aim to eliminate the influence of unique errors of the neural network forecats.

Finally, the quantities λ_q discussed at the beginning of this Section are calculated as slopes of the following regression lines. We take the averaged divergences of perturbed and unperturbed trajectories $s_q(n)$ with respect to "good" sampling points $x(t_i)$, take $d_q(n) = |s_q(n)|$ and consider regression of $\ln(d_q)(n)$ against the forecast number n. Than all regressions are averaged over the "good" neural networks in the committee giving the single regression line for the considered time series. The plots of the averaged values $\ln d(n)$ against the forecast numbers n with corresponding regression lines for the tested time series are shown in Figs. 4-6. The slopes of regression line obviously coincide with $\lambda = \sum_q \lambda_q/N$. We also estimated the standard errors of the final regression slopes. From the first glance these standard errors seem to be most obvious estimations of the Lyapunov exponents errors. However for all considered time series these errors turned out to be one order smaller than $\Delta\lambda$ calculated as standard deviation of Lyapunov exponents given by different neural networks from the committee.

Let us consider in more details the case when the MLE is supposed to be zero (for periodic and quasiperiodic data) and the case when it is small but positive. We propose how to separate these two important cases. Remind that we average the divergences $s_q^i(n) = x_q^f(t_i+n\tau) - \tilde{x}_q^f(t_i+n\tau)$ of perturbed and unperturbed trajectories with respect to all sampling points $x(t_i)$. Here n, $n = 1, ..., H$ is the forecast iteration number and q, $q = 1, ..., N$ is the committee neural network number. Then we construct the linear regression of $d_q(n) = \ln|s_q(n)|$ versus the iteration number n and estimate the MLE λ_q given by the $q-$th neural network as the slope of regression line. Theoretically in the case when $\lambda_q = 0$ the regression determination coefficient (8) is obviously equal to zero. Now $y(n) = d_q(n)$, $\hat{y}(n) = a + \lambda_q n$ and $e(n) = y(n) - \hat{y}(n)$. In practice, the values of $\lambda_q = 0$ turns out to be very small (positive or negative) as well as their mean λ. However the quality of regression can not be estimated due to vanishing values of R^2. In this case we also can calculate MLE or MAE of the mean λ, but they can not be considered as estimation of the maximal Lyapunov exponent errors due to the fact mentioned above. Therefore in this case one has to use a statistical test to check if the maximal Lyapunov exponent significantly differs form zero. The Fisher test [13] implies to consider the Fisher statistics which is linked with the determination coefficient of the considered linear regression as follows $F = (H-2)\frac{R^2}{1-R^2}$, where H is the amount of sampling. In our case H is the horizon of the multistep forecasting. Provided the null hypothesis (slope is equal to zero), the Fisher statistics has the Fisher distribution. Knowing $p-$ levels of the Fisher probability distribution function one can easily show that at $H = 25$ slopes of regression lines (i.e. the estimated values of MLE) do not differ from zero at 10% significance level if corresponding $R^2 < 0.113$. At 5% significance level one should demand that $R^2 < 0.157$.

5 Results of Calculations

The proposed approach was tested on

- two chaotic maps: Henon system and logistic map ,
- periodic time series,
- stochastic time series ,
- EEG signal,
- some transformation of tensotremorogram signal.

The Henon map written as $x_{n+1} = 1 + ax_n^2 + bx_{n-1}$ was taken at $a = -1.4$ and $b = 0.3$. The logistic map $x_{n+1} = r(1-x_n)x_n$ was considered at $r = 4$. As periodic time series the sine was taken. The velocity $v_i = (B_{i+20} - B_i)/20$ with the period 20 of the the fractal Brownian motion (FBM) B_i with the Hurst exponent 0.9 was taken as stochastic time series. We considered velocity to obtain an oscillating series similar to all other considered ones. In EEG signal we took the Cz channel. One of the channels of the tensotremorogram signal [14], [16] was preprocessed by empirical mode decomposition [17], [18] and the second mode was taken as observable. We used empirical mode decomposition to filter noise in the tensotremorogram signal and to single out the component with the aim to detect presence of chaos in the whole signal. Plots of all time series are shown in Figs. 1 - 2.

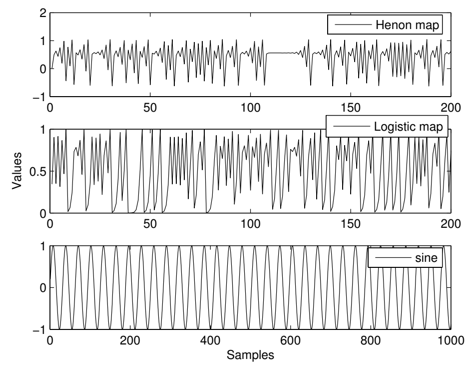

Figure 1: Plots of time series generated by Henon map, logistic map and sine

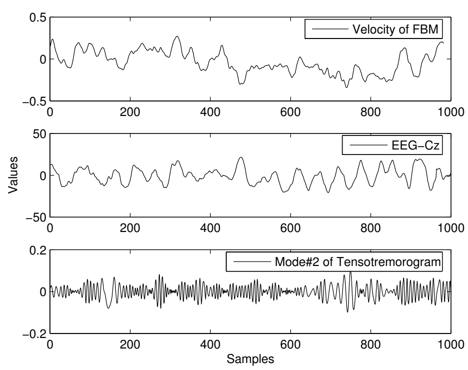

Figure 2: Plots of time series: velocity of FBM, EEG-Cz signal, mode# 2 of tensotremorogram

In Fig.3 we show comparative behaviour of logarithms of absolute values of trajectories divergences for model time series, i.e. Henon and logistic maps, sine and velocity of FBM.

In Figs. 4-6 the plots of $\ln(d(n))$ with regression line are given for logistic map, sine and Cz channel of EEG signal.

In tables below the quantitative characteristics of calculations are given for all time series.

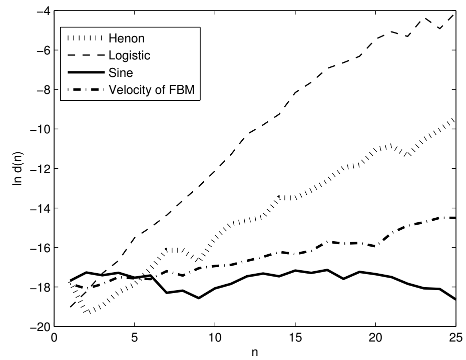

Figure 3: Comparative behaviour of $\ln(d(n))$ plots for stylized time series versus n

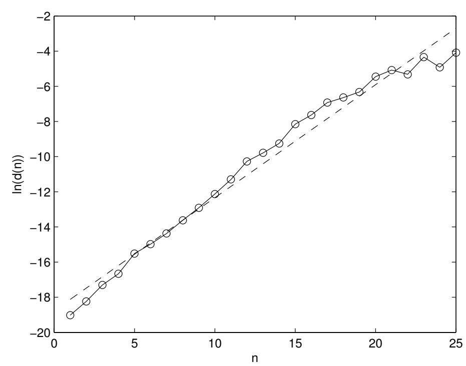

Figure 4: Regression of $\ln(d(n))$ versus n for logistic map

Series	Henon	Logistic	Sine
$\lambda_{theoretical}$	0.42	$\ln 2 \approx 0.69$	0
λ	0.40	0.64	-0.01
$\Delta\lambda$	0.07	0.06	0.06
R^2 regression	0.87	0.92	0.107
Horizon	25	25	25
Length	1000	1000	1000
# of NN	15	30	15
MSE (train)	0.003	0.03	0.0002
R^2(train)	0.99	0.94	0.99

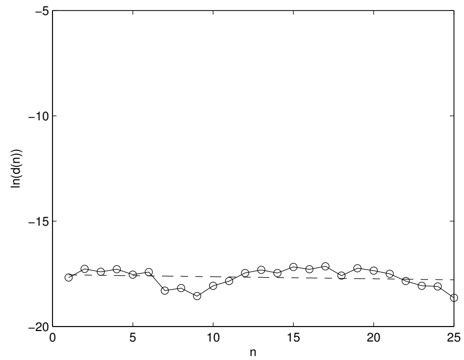

Figure 5: Regression of $\ln(d(n))$ versus n for sine

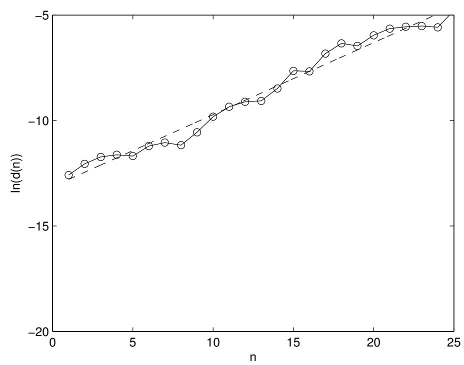

Figure 6: Regression of $\ln(d(n))$ versus n for EEG (Cz channel)

Series	FBM	EEG-Cz	Tensotrem.
$\lambda_{theoretical}$	—	—	—
λ	0.15	0.34	0.30
$\Delta\lambda$	0.11	0.17	0.06
R^2 regression	0.74	0.77	0.79
Horizon	25	25	25
Length	982	1000	1000
# of NN	50	50	30
MSE (train)	0.01	0.003	0.01
R^2(train)	0.94	0.98	0.82

It should be noted that for sine regression $R^2 = 0.107 < 0.113$. Therefore due to the discussion above maximal Lyapunov exponent does not differ from zero at 10% significance level.

It can be seen that for stochastic time series, namely for velocity of FBM, maximal Lyapunov exponent is positive in the range of MLE miscalculation. At the same time statistical characteristics of the neural network training quality (such as R^2 and mean-square error (MSE)) are relatively high. As it has been mentioned above this means that positivity property of the maximal Lyapunov exponent combined with the high statistical characteristics of the neural network training quality, can not be considered as identification of the chaos. Therefore identification of chaos requires additional test and researches. For the Cz channel of EEG signal this has been done in [15]. So the fact that MLE for Cz channel is positive combined with fact of high statistical characteristics of neural networks training quality verify conclusion made in [15] about the chaotic character of this channel of EEG. The similar conclusion we can make for the second mode of empirical mode decomposition of the considered channel of tensotremorogram. Additional studies [14], [16] are in mutual consent with positivity of the maximal Lyapunov exponent calculated here. Despite the quality of neural networks training is less in this case we can assume this signal to be chaotic. But of course it contains strong noise component which reduces the neural networks quality statistics. More detailed discussion of this problem will be done elsewhere.

This work is supported by the RFBR grant N 08-07-12052-ofi.

References

[1] Golovko V.A., Neural Networks Approach to Processing of Chaotic Processes, MIPHI - 2005. VII All Russian Scientific Conference "Neuroinformatics - 2005": Lectures in Neuroinformatics. MIPHI, 2005 - 214 p.

[2] Wolf A., Swift J., Swinney H., Vastano J.,1985, Determining Lyapunov exponents from a time series, Physica D, Vol. 16, pp. 285–301.

[3] Benettin G., Galgani L., Giorgilli A., Strelcyn J.-M, 1980, Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems: A method for computing all of them. Pt. I: Theory. Pt. II: Numerical applications *Meccanica,* Vol. 15, pp. 9–30.

[4] Sano M., Sawada Y., Measurements of the Lyapunov Spectrum from a Chaotic Time Series. Phys. Rev. Lett., 1985, 55, 1082.

[5] Eckman J.P., Oliffson Kamphorst S., Ruelle D., Ciliberto S. Lyapunov Exponents from Time Series. Phys. Rev. A, 1986, 34, 4971.

[6] Rosenstein M.T., Collins J.J., De Luca C.J. Reconstruction Expansion as a Geometry-Based Framework for Choosing Proper Delay Times, Physica D, 1993, 73, 82.

[7] Kantz H. A Robust Method to Estimate the Maximal Lyapunov Exponent of a Time Series. Phys. Lett. A, 1994, 185, 77.

[8] Takens F., 1980, Detecting strange attractors in turbulence. *Lecture Notes in Mathematics* , Vol. 898, Springer-Verlag, Berlin ,pp. 366–381; 1980, and in Dynamical System in Turbulence, Warlock, Eds. D. Rand and L. S. Young.

[9] Kantz H., Schreiber T. Nonlinear time series analysis. - Cambridge: Cambridge University Press, 1997. - 75 pp.

[10] Bezobrazova S.V., Golovko V.A., Neural Networks Model of EEG Segmentation in the Problem of Epileptiform Activity Detection. Lectures in Neuroinformatics; Part. 1. / X All Russian Scientific Conference "Neuroinformatics - 2008", Moscow, 2008. - pp. 11 - 42.

[11] Ezhov A.A., Shumsky S.A. Neurocomputing and its Application in Economics and Business. - Moscow, 1998. - 222 p.

[12] Terekhov S.A. Great Committees of Smart Mashines. Lectures in Neuroinformatics; Part. 2. / IX All Russian Scientific Conference "Neuroinformatics - 2007", Moscow, 2007. - p. 89.

[13] Pismenny D. Lecture Notes in Probability Theory and Mathematical Statistics, Moscow, Airis-Press, 2004, 256 p.

[14] Kuperin Yu.A., Mekler A.A., Dmitrieva L.A., Romanov S.P., Kovalev V.A. Specific Character of the Tensotremorograms EMD-decomposition in Norm and Pathology.// Proceedings of the II International Congress of SNG Neurophisiologysts, Kishinev, Moldova, 29-31 October 2008, Eds. Grigorieva A.I., Sepiashvili R.I., Farduya F.I., Moscow-Kishinev, Medicine- Health, 2008, p.169.

[15] Dmitrieva L.A., Kuperin Yu.A., Mekler A.A., Artificial Neural Networks, Multifractal Analysis in the Study of EEG Signals. //Third International Conference in Cognitive Science : Talks Abstracts: In 2 Vol. Moscow, 20-25 June 2008., Eds. Alexandrov Yu.I., Velichkovsky B.M., Kibrik A.A.,. Kotov E.V., Pechonkova O.E., Svarnik T.V., Chernigovskaya T.V.: Artistically-Publishing Center, 2008. V. 1: - 292 p., pp. 252-254.

[16] Kuperin Yu.A., Dmitrieva L.A., Mekler A.A., Romanov S.P., Kovalev V.A., Kropotov Yu.D., Alexanyan Z.A. Informational Components Extraction from Tensotremorogram Signals by the EMD-decomposition // Regional Informatics -2008 (RI-2008). XI Saint Petersburg International Conference. Saint Petersburg, 22-24 October 2008: Conference Materials. SPb.: SPOISU, 2008. p. 244.

[17] N. E. Huang et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non stationary time series analysis", Proc. Royal Soc. London A, Vol. 454, pp. 903-995, 1998

[18] G. Rilling, P. Flandrin and P. Gonzalvs "On Empirical Mode Decomposition and its algorithms", IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03, Grado (I), June 2003