"Science Stays True Here"
Journal of Mathematics and Statistical Science (ISSN 2411-2518), Vol.6, 85-93 | Science Signpost Publishing

Multifractal Detrended Fluctuation Analysis combined with Singular Spectrum Analysis

Anton Karmatskii

Ural Federal University, 620026 Yekaterinburg, Russia
e-mail: karmatsky.anton@urfu.ru

Abstract. This paper addresses the problem of finding appropriate method for analysis of non-stationary time series with complex trends and possibly with distinct periodic components in power spectrum. A well established Multifractal Detrended Fluctuation Analysis (MDFA) combined with Singular Spectrum Analysis (SSA) used on several artificial examples to demonstrate the capabilities of the method. In this combined method SSA is used for nonparametric periodic components extraction and for adaptive automatic nonparametric trend extraction. As usual, the local fractal characteristics of the signal are studied by utilizing the MFDFA algorithm. It is shown that combined method is capable of accurately extracting fractal features when signal is contaminated with broadband and narrowband noise. The main task that the author focused on when creating the method was the analysis of currents through ion channels in cell membrane.

Keywords: SSA • DFA •nonparametric

1 Introduction

The traditional fractal methods, such as fluctuation analysis (FA) [1], are developed as statistical tools to evaluate the fractal characteristics of the time series, most commonly the scaling exponent. However, FA requires stationarity of time series. To overcome this limitation, a new scale exponent calculation method named detrended fluctuation analysis (DFA) was introduced in [2]. DFA is extensively used to detect the long-range correlation and power-law properties in nonlinear and nonstationary time series, and it is suitable for extraction of precise intrinsic statistical features from the time series by removing external polynomial trends of different orders. Furthermore, detrending helps to avoid the spurious detection of correlations which are artifacts of nonstationary time series.

However DFA deals only with so called monofractals. Monofractal is mainly a description of the overall average of the object of study, with the local characteristics of the signal being insufficiently characterized [3]. Multifractal analysis provides the ability to describe the local characteristics of the signal in detail. By combining multifractal and detrended fluctuation analysis, a multifractal detrended fluctuation analysis (MDFA) algorithm [4] was introduced. Since then, MDFA has been quickly applied in a number of research areas, including turbulence, temperature, stock market time series etc. However, despite its widespread use, it is known that the MDFA has several limitations [3]:

1. When calculating a trend the order of approximating polynomials is not always known a priori and may vary for different lengths of time series. So, several computations with different trends approximations are recommended.
2. The distinct peaks in the power spectrum of the time series (for example, annual harmonic) have a strong influence on the results of the algorithm, so they should be removed by means of appropriate filter.

To solve the above-mentioned problems in present work singular spectrum analysis (SSA) [5, 6] is used. It is a relatively new method of time series analysis that emerged in statistics, as well as in nonlinear dynamics and in classical spectral methods. The method is nonparametric, i.e. does not require an explicit specification of the time series model. SSA can be used for a wide range of tasks: trend or quasi-periodic component detection and extraction, denoising, forecasting, change-point detection.

There are several modifications of original MDFA in terms of trend extraction [3], such as moving average or empirical mode decomposition (EMD), but approach with SSA has several advantages [6]:

- In SSA the type of the extracted trend is determined by the data itself, which reduces the influence of trend features on scaling exponent.
- Ability to control extracted frequencies. Ideally, when dealing with trend extraction in MDFA frequencies $f \leq 1/N$ should be extracted from time series, where N is the number of data points, f is frequency in arbitrary units [3]. In trend extraction by means of SSA such control is part of the algorithm and does not require additional efforts.
- The mathematical theory of SSA guarantees the asymptotic separability of the trend and other components of the time series, such as harmonics or broadband noise.
- For the task of periodic components extraction SSA allow amplitude modulation and, partially, frequency modulation.
- Procedure of trend and periodic components extraction in SSA can be automated [6-8].

Combined algorithm consists of the following steps: first by means of SSA remove narrowband (distinct harmonics) and broadband noise from time series by means of periodic and trend extraction respectively. Then apply MDFA to resulted time series where parametric trend extraction replaced by nonparametric. All required code was written in C++ language.

The remainder of this paper is organized as follows: the principle of SSA, MDFA and combined method are described in Section 2. Section 3 describes test experiments and their results. Finally, conclusions and future directions are given in Section 4.

2 Theory Descriptions

2.1 Singular Spectrum Analysis (SSA)

Suppose we have a time series

$$F = (f_0, f_1, \dots, f_{N-1}) \tag{1}$$

of lenght N, it is assumed that the measurements are carried out at equal time intervals. The first step of the SSA algorithm is the so-called embedding, when a one-dimensional time series is converted to a two-dimensional matrix. The main parameter of embedding is window length 1 < m < N, which is a number of rows in constructed matrix X:

$$\mathrm{X} = \begin{pmatrix} f_0 & f_1 & \cdots & f_{N-m} \\ f_1 & f_2 & & f_{N-m+1} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m-2} & f_{m-1} & \cdots & f_{N-2} \\ f_{m-1} & f_m & & f_{N-1} \end{pmatrix}. \tag{2}$$

Resulted matrix is Hankel matrix [6].

The next step is the Singular Value Decomposition [9] of the matrix X:

$$X = X_1 + X_2 + \cdots + X_d, \tag{3}$$

where d is the rank of the matrix X,

$$X_k = \frac{1}{\sigma_k} \cdot u_k \cdot v_k^T$$

is the k-th component of singular value decomposition, σ_k is the k-th singular value, which is always nonnegative, u_k is the k-th left singular vector, v_k^T is the k-th transposed right singular vector. Set of values (σ_k, u_k, v_k) sometimes called eigentriple. Next, one group different components and reconstruct the time series according to the following rule: average all elements in the sub diagonals:

$$\{x_{pq}, p + q = const\},$$

each such average corresponds to the value of the time series. Grouping of different components of singular value decomposition depends on the problem to solved. For example, for trend extraction all eigentriples in which "low frequencies" predominate (in the left or right singular vectors) are groped to produce trend. For distinct harmonic extraction situation is different: one groups all eigentriples with similar singular values and similar predominated frequencies (in the left or right singular vectors).

SSA algorithm can be automated [7, 8], and automated version is used in this work. An effective SSA algorithm, having practically linear complexity along the length of the time series was made in accordance with papers [10, 11].

2.2 Multifractal Detrended Fluctuation Analysis (MDFA)

MDFA is suitable for nonlinear and nonstationary time series and consists of several steps. Initially, one need to find the cumulative function of the time series defined in (1):

$$Y_i = \sum_{k=0}^{i}(f_k - \overline{f}), \tag{4}$$

$$\overline{f} = \frac{1}{\mathrm{N}}\sum_{j=0}^{N-1} f_j.$$

The next step is to divide the time series into disjoint segments of length s, their number is $N_s = \frac{N}{s}$. Often the length of the segments s does not divide the full length N without a remainder, so, in order not to discard the remainder of the time series, the procedure of division into segments is repeated from the other end of the data. There are $2N_s$ segments in total.

On each of the $2N_s$ segments a local trend is computed and then the variation is calculated:

$$\Psi^2(s,\nu) = \frac{1}{s}\sum_{i=0}^{s-1}(Y_{(\nu-1)\cdot s+i} - y_i^\nu)^2, \tag{5}$$

$$\nu = 1, \dots, N_s,$$

$$\Psi^2(s,\nu) = \frac{1}{s}\sum_{i=0}^{s-1}\left(Y_{N-(\nu-N_s)\cdot s+i} - y_i^\nu\right)^2, \tag{6}$$

$$\nu = N_s + 1, \dots, 2N_s,$$

where y_i^ν is the local trend on a segment ν of length s, which is often calculated by the least squares method, with possible variants including: linear, quadratic, etc.

Further, the resulting variation on each segment is averaged and a fluctuation function is found:

$$\Psi_q(s) = \sqrt[q]{\frac{1}{2N_s}\sum_{\nu=1}^{2N_s}\Psi^2(s,\nu)^{\frac{q}{2}}}, \tag{7}$$

where the index q can take any values that are not equal to zero.

Following [3], we redefine the fluctuation function (we raise it to the power)

$$\Phi_q(s) = (\Psi_q(s))^{\mathrm{q}}. \tag{8}$$

Obviously, the function (8) increases with increasing length and depends on the order of the extracted trend.

In the final step one tests the hypothesis of fractal properties (scaling properties)

$$\Phi_q(s) \approx \mathrm{s}^{\mathrm{h(q)}}. \tag{9}$$

To test the hypothesis (9), a graph is plotted in the double logarithmic scale of the function $\Phi_q(s)$ for each value of q from a certain range. If condition (9) is satisfied, then the graph will have a straight line, the slope of which will give h(q) for a given q.

The function h(q) in this case is called the generalized Hurst exponent. After the corresponding value of h(q) was found for each q, a graph of the dependence of h(q) on q plotted. If the time series is a monofractal, then the graph will have a straight line, with

$$h(q) = H \cdot (1 + q), \tag{10}$$

where H is the ordinary Hurst exponent. For h(q) to be non-linear (so-called multifractality), it is necessary that small and large fluctuations differ in their statistical properties [4].

2.3 Combined MDFA with SSA algorithm

In this paper, it is proposed to replace the parametric method of detrending procedure in the mDFA by nonparametric using the automatic trend extraction by means of SSA. It is also used to automatically extract all periodic components from time series as well as broadband noise before applying MDFA. A full description of the algorithm can be found in [7, 8]. We will briefly describe the algorithm of trend extraction.

We first define disrete Fourier transform and the periodogram of time series (1):

$$\Phi_k = \sum_{n=0}^{N-1} e^{-i\cdot 2\cdot \pi\cdot n\cdot k/N} f_n,$$

$$I_N\left(\frac{k}{N}\right) = \frac{1}{N}\begin{cases} 2|\Phi_k|, if\ 0 < k < N/2 \\ |\Phi_k|, if\ k = 0\ or\ k = N/2 \end{cases}, \tag{10}$$

$$\sum_{n=0}^{N-1} f_n^2 = \sum_{k=0}^{N/2} I_N\left(\frac{k}{N}\right).$$

Then one introduce a measure showing the relative contribution of all frequencies less than the specified one in the periodogram of time series:

$$P_N(\omega) = \sum_{0<\frac{k}{N}<\omega} I_N\left(\frac{k}{N}\right),$$

$$C(\omega_0) = \frac{P_N(\omega)}{P_N(0.5)}, \tag{11}$$

where frequency ω is in arbitrary units and belong to the interval [0; 0.5]. Then all eigentriples (σ_k, u_k, v_k) for which the condition on the left singular vectors is satisfied

$$C(\omega_0) \geq \mathrm{C}_0 \tag{12}$$

will be grouped in the trend. In order to separate colored noise from white noise for ω_0 we use the expression [8]:

$$\omega_0 = \max_{\frac{k}{N}, 0\leq \frac{k}{N} \leq 0.5}\{\frac{k}{N} : I_N(0), I_N\left(\frac{1}{\mathrm{N}}\right), \dots, I_N\left(\frac{\mathrm{k}}{\mathrm{N}}\right) < M_N\}, \tag{13}$$

where M_N is median of the periodogram. The reasoning behind (13) the following: the median of values of the timeseries periodogram gives an estimation of the median of the values of the noise periodogram. So all frequencies $\omega > \omega_0$ correspond to a broadband noise.

The criterion for selecting threshold values for ω_0 and C_0 in condition (12) is described in detail in [7, 8].

3 Experiment

To test the technique, a fractional Brownian motion (with several Hurst exponent H) was generated with total of 10^5 points each. White noise was added to the generated time series with a standard deviation that was two times smaller than for the fractional Brownian motion. In order to simulate narrowband noise two exponentially modulated harmonics with frequencies: 0.05 and 0.2 (in arbitrary units) was also added to the time series.

The resulted time series with H = 0.65 and its periodogram calculated by (10) are presented in Fig. 1. The presence of harmonics and white noise is quite difficult to detect in the time series itself, however, additional patterns are clearly visible on the periodogram: two distinct peaks at frequencies of 0.05 and 0.2, as well as uniform broadband noise at high frequencies.

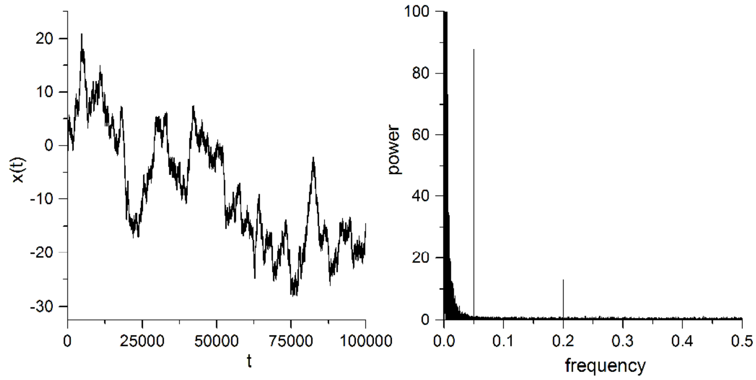

Fig. 1 Sample of fractional Brownian motion with H = 0.65 (left) and its periodogram (right)

We will first apply SSA to extract the periodic components. The result is shown in Fig. 2. It can be seen that despite the amplitude modulation, the extracted modulated harmonics are in good agreement with their true values.

After narrow band noise have been removed, trend extraction with criterion (13) was applied to remove broadband noise. Fig.3 illustrate the influence of ω_0 to trend extraction procedure in SSA. It is clear that when increasing the low frequencies threshold ω_0 in (12), the trend becomes more detailed.

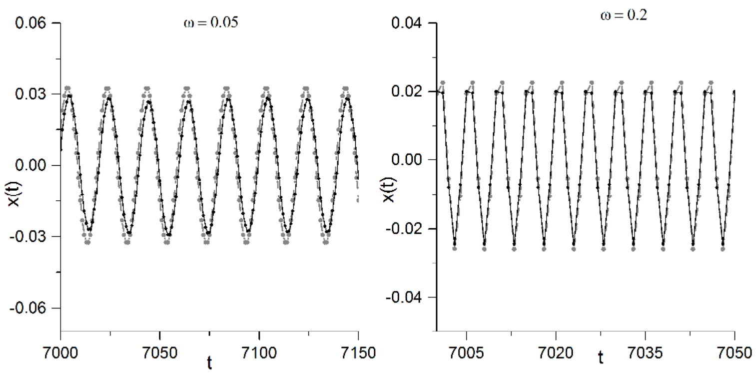

Fig. 2. Extracted periodic components (black) and their real values (grey) with frequencies 0.05 (left) and 0.2 (right).

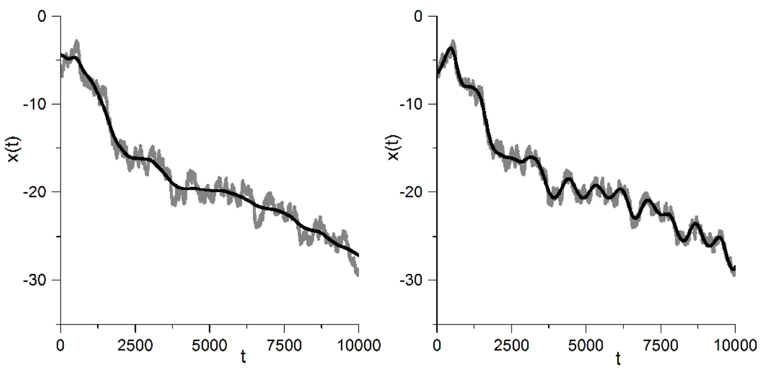

Fig. 3. Example of trend extraction (black line) from Brownian motion (grey line) with $\omega_0 = 10^{-5}$ (left) and $\omega_0 = 10^{-3}$ (right), ω_0 in arbitrary units

When both narrowband and broadband noise have been removed MDFA is applied to the resulted time series. To calculate the fluctuation function and to test the hypothesis (9), the values of q are taken from the interval [-10; 10] with step 0.33. The calculation results are presented in Fig. 4 for several values of q in double logarithmic scale. It can be seen that the points are well approximated by a straight line for all values of the partition lengths. However, for large values of q, the deviation of points from a straight line increases.

For the time series shown in Fig.1 Hurst exponent H was calculated using MDFA and was found to be 0.64 ± 0.02. Similar results were obtained for the case of generated time series with H = 0.25 and H = 0.5, the computed Hurst exponent is equal to 0.24 ± 0.01 and 0.48 ± 0.02 respectively.

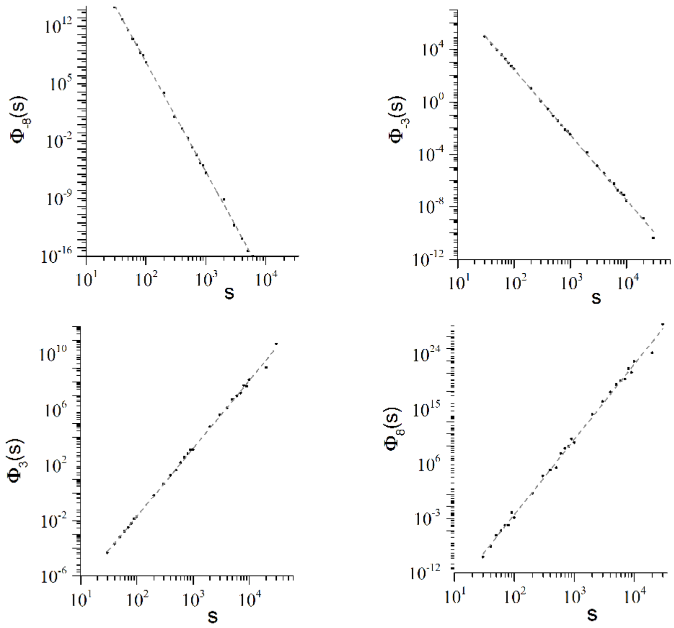

Fig. 4. Fluctuation function for different q and straight line approximation (dashed grey). From left to right: -8, -3 (top); 3, 8 (bottom)

4 Conclusion

A variant of mDFA with an automatic non-parametric trend extraction using SSA has been tested. The capabilities of the method were shown on generated fractional Brownian motion and the method is capable of extracting fractal characteristics in the presence of wideband and narrowband noise. The main task that the author focused on when choosing a technique was the analysis of time series obtained in experiments on ion channels in cells [12].

References

1. Hergarten, S.: Self-organized criticality in Earth systems. Springer (2002). doi: 10.1007/978-3-662-04390-5
2. Peng, C. K., Buldyrev, S. V., Goldberger, A. L. et al.: Long-range correlations in nucleotide sequences. Nature, vol. 356, no. 6363, pp. 168–170 (1992). doi: 10.1038/356168a0
3. Schmitt, F.G., Huang, Y.: Stochastic analysis of scaling time series. Cambridge University Press (2015). doi: 10.1017/CBO9781107705548
4. Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde E. et al.: Multifractal detrended fluctuation analysis of nonstationary time series. Phisica A, V. 316, Issues 1-4, P. 87–114 (2002). doi: 10.1016/S0378-4371(02)01383-3

5. Vautard, R., Ghil, M.: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. Phys. D. V. 35, P.395–424 (1989). doi: 10.1016/0167-2789(89)90077-8
6. Golyandina, N., Nekrutkin, V., Zhigljavsky A.: Analysis of Time Series Structure: SSA and Related Techniques. Chapman and Hall/CRC (2001). doi: 10.1201/9781420035841
7. Alexandrov, T.: Software package for automatic extraction and forecast of additive components of time series in the framework of the Caterpillar-SSA approach: PhD thesis. St.Petersburg State University, 152 p (2006). (In Russian)
8. Alexandrov T.: A Method of Trend Extraction Using Singular Spectrum Analysis. RevStat, V 7, P. 1–22 (2009).
9. Teukolsky, A., Vetterling, W.T., Flannery, B.P., Metcalf M.: Numerical Recipes in Fortran 90 - The Art of Parallel Scientific Computing. Cambridge University Press (1996).
10. Korobeynikov, A.: Computation- and Space-Efficient Implementation of SSA. Stat. Interface, V. 3, P. 357–368 (2010). doi: 10.4310/SII.2010.v3.n3.a9
11. Baglama, J., Reichel, L.: Augmented implicitly restarted Lanczos bidiagonalization method. Sci. Comput. ,V. 27, P. 19–42 (2005). doi: 10.1137/04060593X
12. Neher, E, Sakmann, B.: Single-channel currents recorded from membrane of denervated frog muscle fibres. Nature , 260, P. 799–802 (1976). doi: 10.1038/260799a0