EFFECTIVENESS OF TREND ESTIMATION METHODS

Yu. Cherkas
Kharkiv National University of Radioelectronics,
14, Lenin Ave., Kharkiv 61000 Ukraine
yuriy.cherkas123@gmail.com

Abstract. This paper analyses methods of trend estimation focusing on a process of calculation of their effectiveness. Shown a procedure that helps to choose: “optimal” parameters of trend estimation methods for trends with known characteristics (steepness of increasing or decreasing of trends, a frequency of alternation of regions with positive and negative derivatives etc.). All calculation are shown for data that represents near Earth's asteroids (NEA) semi-major axes distributions. The method is based on running of multiple iterations (in order to provide static stability of results) of a procedure that was designed to find a test signal after adding investigated trend and noise. It allows to calculate a probability of a signal detection after trend removing with using trend estimation methods with different parameters.

Keywords: trend, trend estimation, effectiveness of trend estimation, modeling of effectiveness of trend estimation, near Earth's asteroids semi-major axes distributions.

Introduction

Trend is a long term movement in a series. The problem of trend estimation can be supplied in two main formulations:

1) A series is interpreted as the sum of a systematic component \( u(t) \) (trend) and a random component \( e(t) \) (noise):

\[
z(t) = u(t) + e(t).
\]  

2) A series is interpreted as the sum of a systematic component \( u(t) \), an informational signal \( x(t) \) and a random component \( e(t) \):

\[
z(t) = u(t) + x(t) + e(t).
\]

In the first case, we are interested in detecting of the trend and estimating its parameters. This is required to build a model of the trend in order to investigate the process that generates it or to forecast the further values of the series. In the second case, we are interested in the informational signal, and here it is necessary to get rid of the trend for the subsequent analysis of the residual series.

An effectiveness of detrending with using following method: moving average (Anderson, 1971; Chou, 1975), smoothing splines (De Boor, 2001; Hastie, 1990; Pollock, 1999) and digital signal processing (Boaz, 1997; Lyons, 2010; Smith, 2002) was carried out in the paper. As an object of study for this work are selected near-Earth asteroids (NEA). Sample size at the time of writing was about 11000 orbits (http://neo.jpl.nasa.gov).

Method

The task of identifying and removing the trend is not trivial, since in addition to choosing of detrending method, it is necessary to specify and justify its parameters, which is complicated by the presence of the noise component and the absence of a priori information about an informational signal. An unsuccessful detrending can lead to distortions of the informational signal and its characteristics or can completely suppress it together with the systematic components of the series.

Comparative analysis of detrending methods in relation to the NEA's semi-major axes \( a \) distributions will be carried out according to the following scheme:

1) Formation of a series \( z(a) \) consisting of a standard trend \( u(a) \), test signal \( x(a) \) (a sinusoid with known parameters) and the random component \( e(a) \) (additive white Gaussian noise):
\[ z_1(a) = u(a) + A \sin(2\pi a) \]  
(3)

\[ z_2(a) = u(a) + A \sin(2\pi a) + \varepsilon(a) \]  
(4)

2) Carrying out procedures aimed at finding the test signal in series, with using of detrending methods with different parameters and the subsequent calculation of the periodogram (Brillinger, 1975; Marple, 1987; Terebizh, 1992; Stoica, 1997) of the detrended series. A signal is considered successfully detected, if the difference between the periodogram maximum and the frequency of the test signal is less than or equal to the value of periodogram resolution.

3) Steps 1 and 2 are repeated for different values signal-to-noise ratio (SNR). To ensure the statistical sustainability of the results each configuration of detrending method is simulated for 1000 different realizations of noise with same SNR. That allows to calculate the probability of a successful detection of the signal, which is the main criterion of "optimal" parameters choosing.

Standard trend \( u(a) \) was obtained by plotting the NEA's data distribution along semi-major axis, followed by smoothing the resulting histogram with polynomial of 7 degree (fig. 1). These actions are also trend estimation, but in this case we do not care how well the smoothed polynomial corresponds to the "true" trend, it is only necessary that they have common characteristics. The model \( z_1 \) (3) which contains informational signal with a frequency \( f = 0.2 \) is shown on fig. 2 a). The model \( z_2 \) (4) with additive white Gaussian noise (SNR = 10 dB) is show on fig. 2 b).

Moving average method

The basis of the moving average method was chosen the simplest of this family – the arithmetic moving average. Averaging window width varied from \( m = 3 \) to \( m = 13 \) elements. Contour plots of levels of successful signal detection for different noise power and for different frequencies of informational signal are shown on fig. 3.

On the contour plots ten shades of gray correspond to different levels of probability of successful detection of the test signal. Black color means that probability of detection is \( p < 0.1 \), white – \( p > 0.9 \). Every new gray scale is different from the next by \( \Delta p = 0.1 \). An area of contour plot with \( p > 0.9 \) is called the domain of successful signal detection. From the standpoint of size and uniformity of the domain of successful signal detection acceptable values of averaging window are \( m = 9 \), \( m = 11 \) and \( m = 13 \).
Smoothing with splines

As the basis function of the spline was taken algebraic polynomial of 5 degree. Number of spline knots changed from $l = 3$ to $l = 20$. Plots of levels of successful signal detection for different noise power and for different frequencies of informational signal are shown on fig. 4.

Increasing the number of knots of the spline at the beginning leads to the stabilization of the successful signal detection domain. It becomes almost rectangular shaped after number of spline knots is equal to or
more than $l = 5$. A further change in the number of knots of the spline can adjust a frequency which divides trend components and residual series.

**Digital signal processing**

For digital filtering of trend components of series was chosen elliptic highpass filter with following parameters: attenuation in the stopband $- R_s = 50\text{dB}$, passband ripple $- R_p = 0.1\text{dB}$, the difference between the passband $f_p$ and the stopband $f_s$ edge frequencies $- \Delta f = 0.005$. Corresponding plots of levels of successful signal detection are shown on fig. 5.

The lowest stopband frequency with stabilized domain of successful signal detection is equal to $f_s = 0.02$. Further increasing of stopband frequency leads to a corresponding expansion of suppressed components.

**Discussion**

The method of valuation of trend estimation effectiveness allows to compare and choose "optimal" parameters for different methods of trend estimation. Proposed technique is useful if further frequency analysis of detrended series is supposed. It allows to analyze a distortion of a signal with different frequencies and for various SNR. The calculation of detrending effectiveness of NEA's semi-major axes distributions shows that smoothing splines and digital filtering have the best results according to the chosen criterion (a probability of successful detection of a test signal).

**References**