Decomposition of Time Series – from Statistical to Linear System Approach

[1,2]J.Holcik

[1]Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia
[2]Institute of Biostatistics and Analyses, Masaryk University, Brno, Czech Republic
Email: umerholc@savba.sk

Abstract. *In a field of statistical processing of time series, algorithms used for their decomposition to partial constituent components are ordinarily designed using statistical characteristics of the analysed series. On the other hand, the designed algorithms can be also often characterized by parameters and functions as frequency response, distribution of zeros and poles, impulse response, usually used in case of linear signal processing. The paper describes some classical algorithms for estimation of a time-series drift and analyses their properties using their frequency responses and discusses conditions of their application.*

Keywords: Time Series, Decomposition, Moving Average Model, Linear Frequency Filter

1. Introduction

Standard methodology of statistical time series analysis starts by decomposition of a given time series representing realization of a random variable X_i to several constituent components (e.g. [1]). Depending on either additive or multiplicative model it is theoretically

$$X_i = T_i + Z_i + S_i + R_i \quad \text{or} \quad X_i = T_i \,.\, Z_i \,.\, S_i \,.\, R_i \,, \tag{1}$$

where T_i is a (monotone) function of time called trend, Z_i describes a non-random long term cyclic process, S_i reflects non-random short time periodic seasonal component, and finally R_i is a random noise variable representing all the deviations from the ideal deterministic part of the model. We suppose R_i is a white noise with normal distribution with expectation $E(R_i) = 0$. Sometimes, the variables describing trend T_i and long-term oscillations Z_i are summarized together

$$D_i = T_i + Z_i \,, \tag{2}$$

and the resulting summarized variable D_i is called drift. The basic task of the time series analysis is to determine and separate deterministic slow components T_i, Z_i, or D_i, and periodical seasonal series S_i from random noise R_i.

2. Algorithms for Drift Estimation

Let us assume an additive model now. If we do not deal with models of trend based on functional approximation then the most often used methods for estimation of the slow drift components are based on moving average (MA) approach.

A MA filter is a type of a finite impulse response filter computing a series of weighted averages of consequential segments of the full data series. For weighted MA filter it is valid that

$$y_i^{MA} = \sum_{k=i-L1}^{i+L2} w_k x_i \tag{3}$$

where

$$\sum_{k=i-L1}^{i+L2} w_k = 1 \tag{4}$$

and

$$L = L1 + L2 + 1 \tag{5}$$

defines an order of the MA filter, i.e. length of its impulse response. The coefficients w_k unambiguously defines properties of the filter (in time as well as in frequency domain) and its impact to processed data. If L1 = L2 and the sequence of weight coefficient is symmetrical with respect to central sample then the filter does not introduce any phase distortion. That is why odd number of samples in filter impulse response is usually used in this case. Another reason for preferring impulse response with odd samples is its simpler design.

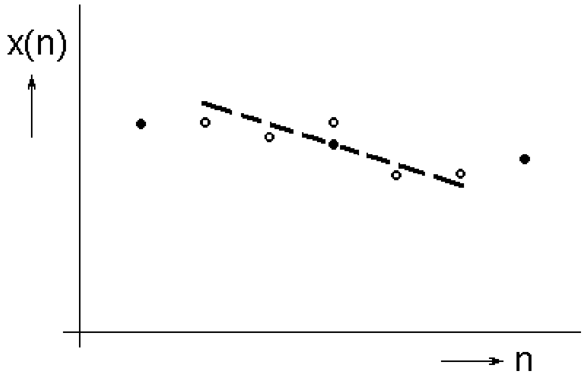

Fig.1 Smoothing of a sequence of five samples by means of a polynomial of the 1st order

In the field of time series statistical processing two basic approaches are used for design of the filter weight coefficients - algorithms based on smoothing by polynomial functions and filters with exponentially descending weights. The former method usually can take advantage of the above mentioned property of the symmetrical sequence of the MA filter weight coefficients, in case of the latter it is not possible, in principle.

3. MA Filters Based on Smoothing by Polynomial Functions

This method (e.g.[2]) is based on an idea that any "reasonable" function can be quite reliably LMS approximated by a polynomial function (Fig.1) of a given order. Then the filtered value is determined as a value of the polynomial function at a position of the substituted sample.

The task necessary to solve before smoothing is to decide which order of the polynomial should be used for the approximation and what the length of the approximated segment is.

Table 1. Weights of MA filters determined by means of polynomial smoothing (partially according to [2])

Length of the segment	Order of the approximating polynomial		
	1st	2nd and 3rd	4th and 5th
3	1/3.(1, 1, 1)	(0, 1, 0)	(0, 1, 0)
5	1/5.(1, 1, 1, 1, 1)	1/35.(-3, 12, 17, …)	(0, 0, 1, ...)
7	1/7.(1, 1, 1, 1, 1, 1,1)	1/21.(-2,3,6,7,...)	1/231.(5,-30,75,131,...)
9	1/9.(1, 1, 1, 1, 1, 1, 1, 1, 1)	1/231.(-21, 14, 39, 54, 59,...)	1/429.(18,-45,-10,60,120,143,..)

Intuitively, we can say that the higher the order of the polynomial is, the broader the frequency pass-band of the filter is. And the longer the segment is, the narrower the frequency pass-band is. It can be often found in statistical literature that the length of the MA segment should correspond to the period of the seasonal component present in the data series. Generally, it can be true. However, relationship between the length of the segment and period of the seasonal component is not so simple.

Tab.1 depicts weights of filter coefficient for different orders of smoothing polynomials and lengths of smoothed data segments. It can be seen that application of the polynomial of the first order results in filter with rectangular impulse response, computing response as uniformly weighted mean value. Such filters correspond to so called Lynn's filters [4], very often used filters in biomedical signal and data processing as ECG signals [5].

In statistical time series processing, data representing analysed processes are often sampled with a month sampling period. Such data (biological, financial, ...) usually have a seasonal component with one year period, it means 12 samples, as well. From the above mentioned recommendation it follows that the length of the filter impulse response should be 12 samples,

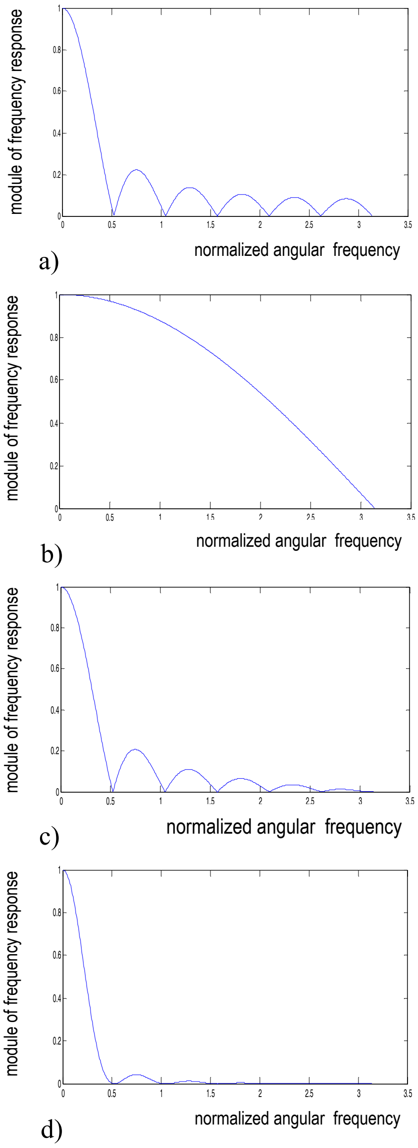

Fig.2 Magnitude frequency responses of the MA filters – a) rectangular impulse response of the length of 12 samples; b) filter with frequency response according to eq.(6); c) filter with frequency response according to eq.(8); d) serial connection of two filters with $H_1(z)$.

as well. Magnitude frequency response $|H_1(z)|$ of such a filter is depicted in Fig.2a. We can see that zero points of the response are exactly at frequencies that correspond to seasonal period and its integer multiples. Unfortunately, the impulse response has even number of samples and that is why a phase frequency response is not precisely linear, even if the deviations from linearity do not look important. To improve a shape of the phase response of the filter, serial connection of the filter smoothing data with a polynomial of the first order with filter having transfer function defined as

$$H_2(z) = \frac{1}{2}(1 + z^{-1}) \text{ . (Fig.2b)} \qquad (6)$$

Impulse response of the resulting filter is

$$g_3(n) = 1/12.\{1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1\} \otimes \otimes \{1/2, 1/2\} = = 1/12.\{0.5, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0.5\}, \qquad (7)$$

which is the most commonly used filter in statistical decomposition of time series. For its transfer function it is valid that (Fig.2c)

$$H_3(z) = \frac{1}{12}(0.5 + z^{-1} + z^{-2} + z^{-3} + z^{-4} + +z^{-5} + z^{-6} + z^{-7} + z^{-8} + z^{-9} + +z^{-10} + z^{-11} + 0.5z^{-12}) \qquad (8)$$

Its zeros remain at the same frequency as zeros of the original filter and attenuation of the side lobes is proportional to that of $H_2(z)$.

To improve attenuation of the higher frequency components (with frequency above the fundamental frequency of the seasonal component) it is possible to use a serial connection of two filters with $H_1(z)$ (Fig.2d).

All the above described filters have zeros at frequencies corresponding to those of periodical seasonal component (integer multiples of 0.52 $rad.sample^{-1}$). It means that the estimated drift is free of any influence of the seasonal component.

In Fig.3 there are magnitude frequency responses of MA filters based on smoothing data by polynomial of higher orders – the 3rd and 5th order. It is obvious that the higher polynomial order represents broader width of the filter pass-band and on contrary longer length of the filter impulse response (with the same polynomial order) decreases filter cut-off frequency. In all the depicted examples, cut-off frequencies of the filter pass-bands are higher than the fundamental frequency of the seasonal component. It means that the seasonal components have to substantially affect outputs of the filters.

4. Filters with Exponential Weights

Formula defining exponential smoothing filter

$$\mathbf{y_i^{EXP}} = (\mathbf{1-b})\mathbf{x_i} + (\mathbf{1-b})\mathbf{b}\mathbf{x_{i-1}} + (\mathbf{1-b})\mathbf{b^2}\mathbf{x_{i-2}} + (\mathbf{1-b})\mathbf{b^3}\mathbf{x_{i-3}} + \cdots \qquad (9)$$

can be rewritten recursively as

$$\mathbf{y_i^{EXP}} = (\mathbf{1-b})\mathbf{x_i} + \mathbf{+b}[(\mathbf{1-b})\mathbf{x_{i-1}} + (\mathbf{1-b})\mathbf{b}\mathbf{x_{i-2}} + (\mathbf{1-b})\mathbf{b^2}\mathbf{x_{i-3}} + \cdots] = (\mathbf{1-b})\mathbf{x_i} + \mathbf{b}.\mathbf{y_{i-1}^{EXP}}$$

This difference equation corresponds to a transfer function

$$H_{EXP}(z) = \frac{1-b}{1-b.z^{-1}} = \frac{a}{1-(1-a).z^{-1}} \qquad (10)$$

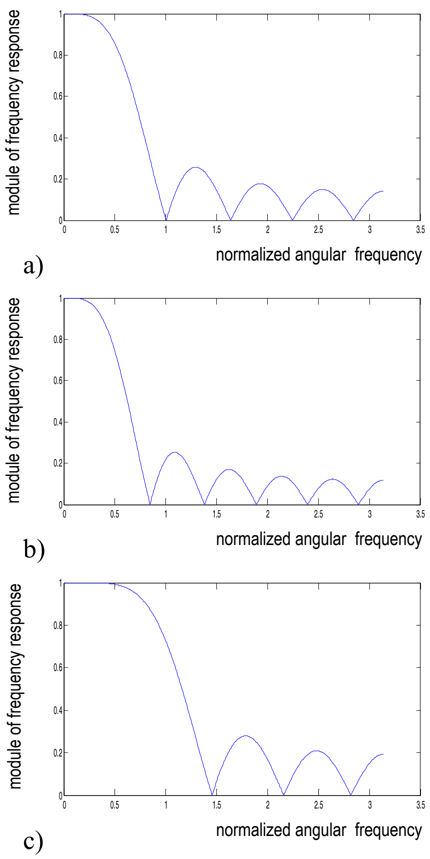

Fig.3 Magnitude frequency responses of filters designed for – a) 3rd order polynomial and 11 sample impulse response; b) 3rd order polynomial and 13 sample impulse response; c) 5th order polynomial and 11 sample impulse response.

where b is so called discount constant and a = 1 – b is a smoothing constant. An example of modular frequency response is in Fig.4. The function is smooth without any zero points. It means that the filter is really not acceptable for processing time series with seasonal component. The width of a pass-band is proportional to the smoothing constant a. The greater the value of the smoothing constant is, the broader is the filter pass-band and the filter is also more stable. It is also necessary to realize that the impulse response is not symmetrical and that is why the phase frequency response of the filter is not linear (Fig.4). That fact can cause a heavy phase distortion in the filtered series.

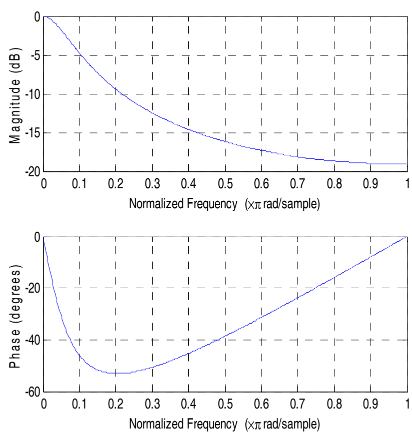

Fig.4 Magnitude and phase frequency response of the filter with exponential weight for smoothing constant a=0.2

5. Conclusions

Frequency responses as well as other ways of description properties of linear algorithms designed using statistical characteristics of the analysed time series can and should be used as an adequate tool for verification of assumed properties of the algorithms and their expected effect to the processed data. Unfortunately, it is not the case in statistical approaches to time series processing.

Acknowledgements

This work was partially granted by the research project No. 2/0210/10 of the VEGA Grant Agency in Slovakia.

References

[1] Falk M. A First Course on Time Series Analysis. Examples with SAS. Chair of Statistics, University of Würzburg, 2006, 240p.

[2] Cipro T. Financial Econometry. Ekopress, Praha 2008, 538p (in Czech).

[3] http://en.wikipedia.org/wiki/Moving_average (Feb 25th, 2011)

[4] Lynn P.A. An Introduction to the Analysis and Processing of Signals. Macmillan, 1989, 277p.

[5] Holcik J, Kozumplik J. Digital Filtering of ECG Signals. Lékař a technika, 17 (6):114-119,1986 (in Czech).