# Spectral analysis of segmented data

S. de Waele and P.M.T. Broersen
Department of Applied Physics, Delft University of Technology
P.O. Box 5046, 2600 GA Delft, The Netherlands
Tel. +31 15 2786419 fax: +31 15 2784263
E-mail: S.deWaele@tn.tudelft.nl

**Abstract**

Time series analysis is reformulated to allow processing of segmented data. This involves the reformulation of parameter estimation and order selection. Parameter estimation for autoregressive (AR) models is done by fitting a single model to all segments simultaneously. Parameter estimation for moving average (MA) and the combined ARMA models can be derived entirely from long autoregressive models. The finite sample theory required for order selection of AR models has be generalized to segments of data. The resulting algorithm can also deal effectively with segments of unequal lenght.

Keywords: segmented data; spectral estimation; time series analysis

## 1 Introduction

Spectral analysis of stationary stochastic signals is used in various fields, e.g. radar [7], fluid dynamics and geophysics [10]. In many situations the duration of one interrupted measurement is submitted to practical limitations. However, it is often possible to obtain several segments of data. The information in the segments can be combined to obtain a more accurate spectral estimate.

With the windowed periodogram, the information can be combined by averaging the periodograms of the individual segments. The generalization of time series analysis to segments of data is less straightforward. However, time series models provide a more accurate spectral estimate than the windowed periodogram [4]. Therefore, time series analysis is used for spectral analysis of segmented data.

An example of the situation where the information in various segments is combined is in the estimation of a clutter model from radar data [7]. Here, a number of segments of equal length is available which can be considered to be independent. Another situation where various segments are obtained is an interrupted measurement of a stationary stochastic process [8]. The interruptions or gaps can consist of one or several observations. Here, the segments are of unequal length. Moreover, if a gap between two segments is small the segments will be correlated.

In time series analysis, a signal $x$ is modeled as an ARMA$(p, q)$-process. This model is determined in two stages: parameter estimation for model orders 1,..., $L_{max}$ and selection of model order and type. With the ARMAsel algorithm both parameter estimation and order/type selection are done without using prior information about the signal properties [4]. The power spectrum and the autocorrelation function are calculated from the estimated time series model. In this paper the ARMAsel algorithm is generalized to allow processing of segmented of data.

The order selection for AR models requires a generalization of the finite sample formulae, which are already known for a single segment [5]. For MA and ARMA order selection, the residual variance to be used in a selection criterion is calculated by fitting the model to all segments taken together. The described algorithm also allows the analysis of segments of unequal length.

## 2 Parameter estimation

The usual way of combining the information contained in segments is to take the average of the models estimated from the individual segments. This will reduce the variance of the estimate, but does not reduce the bias. The bias can be reduced by fitting a single model to all segments taken together [6].

The Burg algorithm is an accurate estimator for AR-parameters [1]. With the Burg algorithm, reflection coefficients $k$ are estimated directly with an iterative algorithm [9]. More specifically, an AR$(p)$ model is fitted to the data with the restriction that the best AR$(p-1)$-model of the AR$(p)$-model is fixed to the AR$(p-1)$ model found in the previous step. This results in the estimation of a reflection coefficient $k_p$ in each iteration step while all previous reflection coefficients $k_1, ..., k_{p-1}$ are fixed. The resulting AR$(p)$ model is guaranteed to be stationary. This algorithm is generalized to segments by estimating the reflection coefficient from all segments taken together:

$$\hat{k}_p = -\frac{\sum_s \left\langle \bar{v}^s[p], \bar{w}^s[p] \right\rangle}{\sum_s \frac{1}{2}\left( \left\| \bar{v}^s[p] \right\|^2 + \left\| \bar{w}^s[p] \right\|^2 \right)}. \qquad (1)$$

where $\bar{v}^s[p]$ and $\bar{w}^s[p]$ are the forward and backward residuals for segment $s$ and order $p$. A comparison with a variety of averaging methods showed that this Burg algorithm is the best way of estimating a AR-model from segmented data [6].

This algorithm is particularly useful for estimation from segments of unequal length. Suppose an AR model of order 20 is to be estimated from 10 segments of 10 observations and 1 segment of 100 observations. Then, reflection coefficients $k_1$ up to $k_9$ can be estimated using all segments. Reflection coefficients $k_{10}$ up to $k_{20}$ are estimated from the remaining segment of 100 observations. This efficient usage of all segments is due to the iterative estimation in the Burg algorithm. This kind of efficiency cannot be achieved with the Yule-Walker or least squares estimators.

MA models can be estimated from an intermediate AR model using the Durbin estimator. This estimator has been made efficient by using the correct order for the intermediate AR model [3]. Similarly, ARMA estimation can be done by fitting an ARMA model to an intermediate AR-model. By selecting the most accurate estimate from several linear reduced-statistics estimators the accuracy is improved [2].

## 3 Order selection

For MA and ARMA, an asymptotic order selection criterion can be used. For AR models, order selection is done using finite sample theory [1]. The order selection criterion which is used, the Combined Information Criterion CIC, requires an expression $v_i$ for the variance of reflection coefficients $k_i$ estimated in white noise. The $v_i$ for segments of data are given by:

$$v_i = var\,(k_i) = \frac{1}{\sum_{s=1}^{S} \max\left((N_s - i)\,, 0\right) + 1}, \qquad (2)$$

where $S$ is the number of segments and $N_s$ is the number of observations in segment $s$. Asymptotically, $v_i$ reduces to $\frac{1}{\sum N_s}$. It can be used for order selection in independent segments. Surprisingly, it can also be used for segments which originate from a single long signal with missing observations. At first sight, one would expect that the variance expression would be different for this situation, as the segments are correlated. However, when an AR-model of sufficiently high order is estimated, the residuals become white noise. As a result, the residuals in the various segments will be independent, even when the original segments are correlated. This has been tested in a simulation experiment of an AR(2)-process with $k_1 = -0.9$, $k_2 = 0.8$. A single signal of 200 observations has been split up into 10 segment of 10 observations and 1 segment of 100 observations to create heavily correlated segments. For orders 1 and 2 the residuals are not white noise, because there are still significant parameter to be estimated. This results in a deviation between the $v_i$ and the actual variance of estimated $\hat{k}_i$ (see figure 1). These deviations do not adversely influence the quality of order selection. For $p > 2$, the variance of estimated reflection coefficients is accurately described by equation 2.

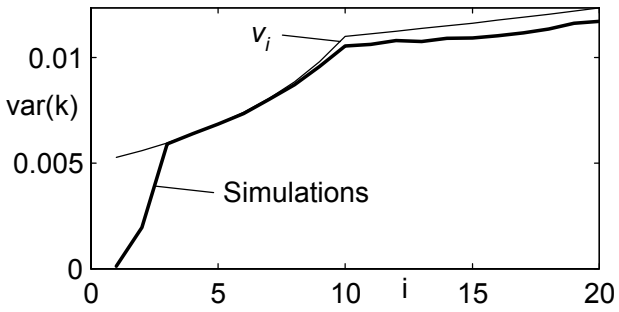


**Figure 1:** Variance of reflection coefficients $k$ estimated from correlated segments of an AR(2) process. Simulation results ($10^5$ runs) are compared to the $v_i$.

## References


[1] P. M. T. Broersen. The ABC of autoregressive order selection criteria. In *Preprints Sysid '97 Conf.*, pages 231–236, Kitakyushu, Japan, July 1997.

[2] P. M. T. Broersen and S. de Waele. Improving the efficiency of reduced statistics ARMA estimators. In *Proceedings of SYSID 2000*, pages 1–6, Santa Barbara, USA, June 2000.

[3] P.M.T. Broersen. Autoregressive model orders for durbin's MA and ARMA estimators. *IEEE Trans. Signal Proc.*, 48(8):2454–2457, august 2000.

[4] P.M.T. Broersen. Facts and fiction in spectral analysis. *IEEE Trans Instrumentation Measurement*, 49(4), august 2000.

[5] P.M.T. Broersen and S. de Waele. Finite sample criteria for autoregressive order selection. *acccepted for publication in IEEE Trans. Signal Processing.*

[6] S. De Waele and P. M. T. Broersen. The burg algorithm for segments. *IEEE Transactions on Signal Processing*, 48(10), october 2000.

[7] S. Haykin, B. W. Currie, and S. B. Kesler. Maximum-entropy spectral analysis of radar clutter. In *Proceedings of the IEEE*, volume 70, pages 953–926, september 1982.

[8] R. Pintelon and J. Schoukens. Frequency domain system identification with missing data. *IEEE Trans. Automatic Control*, 45(2):364–369, february 2000.

[9] P. Stoica and R. L. Moses. *Introduction to Spectral Analysis*. Prentice Hall, Upper Saddle River, 1997.

[10] P. Yuou, E. Baert, and M.F. Loutre. Spectral analysis of climate data. *Survey in Geophysics*, 17:619–663, 1996.