Automatic Spectral Analysis With Time Series Models

Piet M. T. Broersen

***Abstract*—The increased computational speed and developments in the robustness of algorithms have created the possibility to identify automatically a well-fitting time series model for stochastic data. It is possible to compute more than 500 models and to select only one, which certainly is one of the better models, if not the very best. That model characterizes the spectral density of the data. Time series models are excellent for random data if the model type and the model order are known. For unknown data characteristics, a large number of candidate models have to be computed. This necessarily includes too low or too high model orders and models of the wrong types, thus requiring robust estimation methods. The computer selects a model order for each of the three model types. From those three, the model type with the smallest expectation of the prediction error is selected. That unique selected model includes precisely the statistically significant details that are present in the data.**

***Index Terms*—Covariance estimation, identification, order selection, parametric model, spectral estimation.**

I. Introduction

TIME series analysis uses estimated models to compute the spectral density and the covariance function of stochastic observations. This parametric approach is a modern perspective for the nonparametric approach with windowed periodograms [1]. Spectral analysis with tapered and windowed periodograms has been the main practical tool for a long time. It can be described as squared Fourier transforms of data, or as a transform of biased lagged-product covariance estimates [1]. Both descriptions can lead to the same spectral estimate.

Time series models are to be preferred for spectral estimation if the true model type and model order of the process under investigation are known, because only $p-q$ sample covariances of an $\mathrm{ARMA}(p,q)$ process are efficiently estimated with lagged products [2]. Therefore, the covariance estimates for higher lags are not efficient; often, they may be very inefficient. However, periodogram estimates for the spectral density are, at least partially, based on those high lag covariance estimates. Using only the few efficient covariance estimates certainly produces heavily biased spectral estimates.

The increased speed of the computations has recently created the possibility to compute a large number of candidate time series models for one given set of data. The preferred model type and model order for the data are automatically selected with statistical criteria [3], [4]. From that single selected model, the covariance and the spectral density can be computed. The accuracy of the parametric spectrum is typically better than the best of all possible periodogram estimates, and the Cramér-Rao lower bound is often approached in simulations [4]–[6].

So far, time series identification, or the selection of the model type, has not been discussed in the literature. Recently, an automatic selection algorithm for the model type has been proposed [3]; the single selected AR, MA, or ARMA model is denoted the ARMAsel model. This paper describes some historical developments that finally resulted in robust algorithms that can be used in the automatic identification of the model type and order.

Manuscript received December 19, 2001; revised January 8, 2002.

P. M. T. Broersen is with the Department of Applied Physics, Delft University of Technology, Delft, The Netherlands (e-mail: broersen@tn.tudelft.nl).

Publisher Item Identifier S 0018-9456(02)02917-0.

II. Historical Overview

Spectral estimation has a long history where the progress has been influenced alternately by theoretical and by computational developments. Mainly *stationary* stochastic processes are studied theoretically. This seems to be a severe mathematical restriction for measured random data. In practice, however, the definition of stationarity can be treated very loosely. For example, data like speech often can be considered as stationary enough over small intervals, and their spectra have useful interpretations. A clear computational influence was the use of the FFT algorithm of Cooley and Tukey for Fourier transforms [7]. The reduced computer effort enabled the routine analysis of extensive sets of data with periodogram analysis. Therefore, nonparametric spectral analysis with tapered and windowed periodograms has been the main practical tool for spectral analysis for a long time.

Some historical developments in spectral estimation show the combined growth of both parametric and nonparametric methods. More than a century ago, Schuster [8] used periodogram analysis to find hidden periodicities. In 1927, Yule published an article about autoregressive models [9]. Throwing peas on the pendulum, thus giving a physical introduction to autoregressive modeling, supposedly causes deviations from a pure harmonic motion in a pendulum. The first description of real data with moving average models is attributed to Slutsky; it was translated in 1937 but written ten years earlier [10]. Time and frequency domain considerations were united for stochastic processes by the independent contributions of Wiener in 1930 [11] and Khintchine in 1934 [12].

Maximum likelihood is a reliable principle to derive efficient estimators in ordinary linear regression problems. Mann and Wald [13] proved in 1943 that it is ***for large N*** justified to use the same AR data both as dependent and as independent variables in regression theory and to apply the maximum likelihood principle. This was a reason for Whittle to state in 1953: "*it is by no means obvious that these* (ML) *properties are conserved when*

the sample variates are no longer independent, e.g., in the case of a time series" [14]. He also showed that using only the first two sample covariances to estimate the MA(1) model is very inefficient, although all higher true covariances are zero. The efficiency loss is a factor 3.6 for $b_1 = 0.5$ and more than 100 000 for $b_1 > 0.98$. Under rather mild conditions, the maximum likelihood estimator of a function of a stochastic variable is equal to the function of the maximum likelihood estimate of the variable itself [15]. Hence, spectra of the MA(1) process, obtained with lagged-product covariance estimates are far from efficient. The covariance can be estimated much more accurately from the MA(1) model [2], [14]. Arato [16] showed in 1961 that the first $p+1$ lagged products are asymptotically a sufficient estimator for an $\mathrm{AR}(p)$ model. This is no longer true for other types of processes. In fact, $0 \leq k \leq p-q$ sample covariances are asymptotically efficient in an $\mathrm{ARMA}(p,q)$ process [2]. Hence, lagged- products are not efficient covariance estimates and the ***periodogram*** can be nothing but a "***quick and dirty***" estimator of the spectral density.

Maximum likelihood estimation is a nonlinear operation for MA and for ARMA. That gives problems with convergence and invertibility of models, especially in small samples. Different computationally efficient algorithms have been derived by using approximations to the likelihood. Durbin introduced in 1959 an algorithm for MA estimation from a long estimated AR model [17]. A year later, Durbin [18] used the long AR model to reconstruct estimated residuals to be used in ARMA estimation and an additional alternating update of the MA and the AR parameters. Burg [19] described in 1967 a very robust AR estimator that estimates one reflection coefficient at a time from forward and backward residuals. Meanwhile, after 1965 [6], the FFT caused a revival of periodogram-based spectral estimation. The book of Box and Jenkins [20] in 1970 showed how time series models could be estimated from practical data. Pioneering work on order selection has been done by Akaike, who, in 1974, introduced the celebrated selection criterion AIC [21]. Parzen [22] discussed the relation between time series models and prediction. Also, the physical and econometrical modeling of a truly $\mathrm{AR}(\infty)$ process with finite order AR approximations is treated. Priestley [1] described in 1981 the situation of the parametric and nonparametric spectral estimators in a mathematically accessible style. Kay and Marple [23] conclude after an extensive survey of different algorithms for time series models that "when the model is an accurate representation of the data, spectral estimates can be obtained whose performance exceeds that of the classical periodogram." In other words, if model type and model order are known *a priori*, time series models give the best solution. It would be until about 2000 before a successful and robust attempt was reported to select the model type and order from stochastic data with unknown characteristics [3]. The key for this solution was to incorporate finite sample selection criteria in the software.

III. Facts About Periodograms

Periodograms can be characterized as "***quick and dirty***" for stochastic data. Despite all efforts to construct efficient spectral estimators, the best or optimal spectral window can only be determined if the true spectrum is known a priori [1]. The "quick" part of periodograms is due to the application of the FFT [6]. The inevitable characterization "dirty" has more or less coherent reasons.

- Each periodogram can be considered as the transform of a finite length of estimated covariances and as such is the MA model. Comparison of this periodogram with directly estimated MA models shows that the accuracy of the periodogram is less than the accuracy of estimated and selected MA models [5].
- Only $0 \leq k \leq p-q$ estimated sample covariances are asymptotically efficient in an $\mathrm{ARMA}(p,q)$ process [2]. More than $p-q$ estimated covariances are always required to estimate the periodogram. The inefficient extra sample covariances can only produce inefficient periodogram estimates for the spectral density.
- In order to obtain a positive semi-definite covariance estimate, the mean-lagged-product estimator has a triangular bias, using $1/N \sum x_n x_{n+k}$ as estimator for $N-k$ lagged products [1].
- FFT computations treat all data as periodical and produce artifacts by treating the first and the last observation as neighbors.
- To reduce the triangular bias and this latter artifact, data have to be tapered or multiplied by a data window before the Fourier transform is computed, thus introducing a distortion at the ends of the data.
- The Fourier transform of a stationary stochastic process does not exist, because the infinite sum of the absolute values of a realization of a stochastic process is not finite [1]. Hence, taking more observations does not converge to a better approximation, and statistical rules about what is a better estimate cannot be derived from the asymptotical properties of that nonexisting transform of infinite length.
- *A priori* knowledge of the true spectrum is required to determine the optimal variance reduction window.

Those restrictions show that windowed and tapered periodograms cannot be accurate estimators of the spectral density of stochastic processes. They remain ***quick and dirty***. Periodograms may give very good spectra for periodic processes, but not for stochastic.

IV. Time Series Definitions

Three types of time series models can be distinguished, autoregressive or AR, moving average or MA and the combined ARMA type. An $\mathrm{ARMA}(p,q)$ process can be written as [1]

$$x_n + a_1 x_{n-1} + \cdots + a_p x_{n-p} = \varepsilon_n + b_1 \varepsilon_{n-1} + \cdots + b_q \varepsilon_{n-q} \quad (1)$$

where ε_n is a purely random process, thus a sequence of independent identically distributed stochastic variables with zero mean and variance σ_ε^2. This process is purely AR for $q = 0$ and MA for $p = 0$. Any stationary stochastic process with a continuous spectral density can be written as a unique $\mathrm{AR}(\infty)$ or $\mathrm{MA}(\infty)$ process. The roots of

$$A(z) = 1 + a_1 z^{-1} + \cdots + a_p z^{-p} \quad (2)$$

are denoted the poles of the $\mathrm{ARMA}(p, q)$ process, and the roots of

$$B(z) = 1 + b_1 z^{-1} + \cdots + b_q z^{-q}$$

are the zeros. Processes and models are called stationary if all poles are strictly within the unit circle, and they are invertible if all zeros are within the unit circle. The power spectrum $h(\omega)$ of the $\mathrm{ARMA}(p, q)$ process is for frequencies between $-\pi$ and π given by [1]

$$h(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{\left|1 + \sum_{i=1}^{q} b_i \exp(-j\omega i)\right|^2}{\left|1 + \sum_{i=1}^{p} a_i \exp(-j\omega i)\right|^2}. \tag{3}$$

The covariance function of the process is defined as the inverse continuous Fourier transform of (3). It can be approximated as the inverse discrete Fourier transform of (3) by sampling $h(\omega)$. This sampling causes aliasing in the time domain if the covariance function is longer than half the number of samples used. However, direct methods to exactly derive the covariance from the $\mathrm{ARMA}(p, q)$ parameters exist and are given in the time series literature [1], [24]. They use the well-known Yule-Walker relations for the AR covariance and lagged parameter products for the MA part [1], [23].

A measure for the difference between two processes or between two estimated models is the model error ME. In simulations, it is a measure for the accuracy of estimated models by comparing the estimated model with the true process. The ME is defined as a scaled transformation of the one step ahead squared error of prediction PE. With a process given by (1) and $\hat{A}(z)$, $\hat{B}(z)$ denoting the estimated model, ME is defined as [24]

$$\mathrm{ME} = \mathrm{ME}\left(\frac{\hat{B}(z)}{\hat{A}(z)}, \frac{B(z)}{A(z)}\right) = N\left(\frac{\mathrm{PE}}{\sigma_\varepsilon^2} - 1\right) \tag{4}$$

where N is the number of observations used in estimating the model parameters. For unbiased models, the asymptotical expectation of ME is the number of estimated parameters.

V. Requirements for Time Series Algorithms

Some remarks and requirements are given for algorithms that can automatically estimate parameters of many AR, MA, and ARMA models for all sorts of data and select the best model type and order for stochastic data.

- It must be possible to find a good spectral estimate without the requirement that users are aware of details of the method or have to make important choices; of course, they must be able to interfere if they want.
- An algorithm is necessarily a compromise; it cannot be the very best for all possible kinds of random data, but it must always give acceptable results, not too far from the Cramér-Rao boundary.
- The algorithm must be capable of computing very high model orders, higher than 1000 for p and q in (1).
- Only *stationary* and *invertible* models are acceptable. Any method that ***can*** produce roots of estimated polynomials outside the unit circle is not suitable; so far, this includes all not constrained least squares algorithms.
- In finite samples, maximum likelihood estimation is less accurate than some robust methods [25].
- Many algorithms converge asymptotically to the same result, but have a different finite sample behavior; therefore, preferences are mostly based on finite samples and cannot be based only on the asymptotical theory.
- Adding p zeros before and after the data causes the result that several popular $\mathrm{AR}(p)$ estimation methods yield *exactly* the same parameters. Hence, the asymptotical results for ever increasing sample size N will be the same. Burg's method [19] is chosen as the best for AR estimation, with small bias and small finite sample variance [4]. The Yule-Walker method has too much bias [26]; least squares methods cannot guarantee stationarity and have a greater variance; and, finally, full ML solutions are less accurate in finite samples.
- It is necessary to use a finite sample criterion for AR order selection if the maximum order is greater than $0.1N$ [4]; such high AR orders are often necessary as intermediate models for MA or ARMA estimation [27].
- Durbin's MA method has for a long time been considered as inaccurate because a wrong choice for the long AR order was used; but, the recently described new finite sample choice for the intermediate AR order [27] improves the accuracy until close to the Cramér-Rao bound for many examples. So far, Durbin's method is the only method known to the author that can calculate MA models of very high orders, say MA(1000) and at the same time guarantee invertibility.
- Also, Durbin's ARMA methods were not accurate enough until the discovery of a better choice for the intermediate AR order [27]. The estimated models are always invertible and stationary for one specific variant of Durbin's method [28], which allows very high ARMA orders.
- Asymptotic order selection criteria can be used for MA and ARMA estimation, but the penalty factor for additional parameters should be 3 instead of the penalty 2 used in AIC. The higher penalty gives a better compromise between underfit with too few parameters and overfit with too many [29]. The use of consistent order selection criteria is discouraged because the underfit error can become large, especially if consistent criteria are applied to data of the wrong type where the "true" order will be infinite.
- All algorithms must perform well on all types of data. MA and ARMA algorithms must also give reliable results for AR data. In the selection stage, it will turn out that those models give a poor fit to AR data and are not selected at the end.
- For reduction of the computing time, the ARMA estimation has been limited to only the hierarchical $\mathrm{ARMA}(r, r-1)$ models. This also improves the quality of order selection because the selection becomes easier and better if less competitive candidates are available.
- A measure for the quality of models is necessary for a qualification "better" and "worse." The best model has the smallest prediction error or ME and is at the same time the model with a white residual spectrum. This means that all frequencies are equally important in estimating that

model. In other words, that model optimizes the relative error in the frequency domain, which is the same as the absolute error in the logarithm of the spectrum [24].

- Problems can always be expected if the length of the measured data is too short. This means shorter than the impulse response, shorter than the correlation length or shorter than the inverse correlation length, but those problems are probably present in all general applicable methods.
- The algorithm must be able to select type and order, but also to calculate the $\mathrm{AR}(p)$, $\mathrm{MA}(q)$ or $\mathrm{ARMA}(p,q)$ model with prescribed p and/or q.

VI. Robust Time Series Algorithms

A. AR Estimation

This model type is the backbone of time series analysis in practice. Burg's method, also denoted maximum entropy, estimates the reflection coefficients [19], [23], thus making sure that the model will be stationary, with all roots of $\hat{A}(z)$ within the unit circle. Asymptotic AR order selection criteria can give wrong orders if the candidate orders are higher than $0.1N$. The finite sample criterion $\mathrm{CIC}(p)$ is defined as [4]

$$\mathrm{CIC}(p) = \ln\{\mathrm{RES}(p)\} + \max\left[\prod_{i=0}^{p} \frac{1+\frac{1}{N+1-i}}{1-\frac{1}{N+1-i}} - 1, 3\sum_{i=0}^{p}\frac{1}{N+1-i}\right]. \quad (5)$$

The order p with the smallest value $\mathrm{CIC}(p)$ is selected. CIC uses a compromise between the finite sample estimator for the Kullback–Leibler information [30] and the optimal asymptotic penalty factor 3 [4], [29].

B. MA Estimation

Durbin's method for MA estimation guarantees invertibility with all zeros inside the unit circle [17]. Theoretically, the $\mathrm{MA}(q)$ model is equivalent with the $\mathrm{AR}(\infty)$ model, by using $B(z) = 1/A_{\infty}(z)$. Durbin's method uses the *estimated* parameters of a long AR model to approximate the MA model. Of course, the order of that long AR model $A^{\mathrm{long}}(z)$ has to be finite in estimation. The true MA process and the approximating long AR representation are given by

$$\begin{aligned} x_n &= B(z)\varepsilon_n \\ A^{\mathrm{long}}(z)x_n &\approx \varepsilon_n. \end{aligned} \quad (6)$$

Substitution of x_n in the second equation yields the result

$$A^{\mathrm{long}}(z)\hat{B}(z) \approx 1. \quad (7)$$

The MA parameters $\hat{B}(z)$ are computed by minimizing the sum of squares of ***all*** powers of z^{-1} in (7), from 0 to $\dim(A_{\mathrm{long}})+q$; the coefficient of z^{-n} in this series expansion becomes

$$a_n^{\mathrm{long}} + a_{n-1}^{\mathrm{long}}\hat{b}_1 + \cdots + a_{n-q}^{\mathrm{long}}\hat{b}_q \approx 0 \quad (8)$$

where the estimated parameter b_0 is always taken to be 1. The performance of Durbin's algorithm has been improved by selecting a better order for the intermediate AR model [27]. That order is chosen, with a sliding window algorithm, as twice the AR order which is selected for the data with $\mathrm{CIC}(p)$ plus the number of MA parameters to be estimated. The MA order q is selected with the asymptotical criterion $\mathrm{GIC}(q,3)$, defined as [4]

$$\mathrm{GIC}(q,3) = \ln\{\mathrm{RES}(q)\} + \frac{3q}{N}. \quad (9)$$

$\mathrm{RES}(q)$ is the residual variance that can be computed by filtering the original N observations of the process with the inverse of the estimated MA model $\hat{B}(z)$. Applying the inverse of the MA model is the same as using an AR filter with the parameters of that MA model.

C. ARMA Estimation

ARMA models can be computed with Durbin's methods [18]. The first ARMA method of Durbin uses reconstructed residuals $\hat{\varepsilon}_n$ and previous observations x_n as regressors in a least squares solution. The ARMA process is given by

$$\begin{aligned} A(z)x_n &= B(z)\varepsilon_n \\ A^{\mathrm{long}}(z)x_n &= \hat{\varepsilon}_n. \end{aligned} \quad (10)$$

The residuals $\hat{\varepsilon}_n$ are reconstructed from x_n by using a long AR model $A^{\mathrm{long}}(z)$. If the order of $A^{\mathrm{long}}(z)$ would be infinite and the parameters would exactly be the series expansion of $A(z)/B(z)$, the residuals $\hat{\varepsilon}_n$ could be perfect. Durbin's first method estimates the parameters by minimizing

$$\sum_{n=\max(p,q)+1}^{N} \{x_n + \hat{a}_1 x_{n-1} + \cdots + \hat{a}_p x_{n-p} - \hat{\varepsilon}_n - \hat{b}_1\hat{\varepsilon}_{n-1} - \cdots - \hat{b}_q\hat{\varepsilon}_{n-q}\}^2. \quad (11)$$

This ARMA solution is not efficient, and not guaranteed to be stable and invertible. Using the AR parameters of Durbin's first ARMA method as initial conditions, sequential updates of the MA parameters and AR parameters can be made with Durbin's second method. For updating the MA parameters, an adaptation of Durbin's MA method is used. Substitution of x_n from the first equation into the second part of (10) and replacing true by estimated parameters give

$$\frac{A^{\mathrm{long}}(z)}{\hat{A}(z)}\hat{B}(z) \approx 1. \quad (12)$$

A sliding window choice for the order of the intermediate AR model $A^{\mathrm{long}}(z)$ has been described for $\mathrm{ARMA}(r, r-1)$ models [27]. It is three times the AR order as selected with $\mathrm{CIC}(p)$ plus the number of parameters that has to be estimated. By using the long AR model divided by the initially estimated $\hat{A}(z)$ from (9), the MA parameters $\hat{B}(z)$ are estimated from the quotient $A^{\mathrm{long}}(z)/\hat{A}(z)$ in (12), like they were in (8) from $A^{\mathrm{long}}(z)$. Having obtained this new estimate $\hat{B}(z)$ for the MA parameters, the initial estimate $\hat{A}(z)$ obtained with (11) is updated with a new representation of (12)

$$\hat{A}(z) \approx A^{\mathrm{long}}(z)\hat{B}(z). \quad (13)$$

The solution can be written as

$$\hat{A}(z) = \arg\min_{\tilde{A}(z)} \left\{ \mathrm{ME}\left(\frac{\hat{B}(z)}{\tilde{A}(z)}, \frac{1}{A^{\mathrm{long}}(z)} \right) \right\}. \qquad (14)$$

This solution for the p parameters in $\hat{A}(z)$ can also be interpreted as equating the covariance function of $\hat{A}(z)$ to the first p terms of the covariance generated by the right-hand side product of (13), or to the first p reflection coefficients of that product. The Yule–Walker relations [23] can be used to calculate p AR parameters from p covariances, using the reflection coefficients. The accuracy of the second method of Durbin is very much dependent on the proper choice of the order of $A^{\mathrm{long}}(z)$. However, taking the long order too high is much less detrimental than taking the order too low. After computing the model parameters, the residuals can be computed by filtering the data with the inverse of the estimated ARMA model. Like in (9), the penalty factor 3 can also be used for selection of the $2r-1$ parameters of the $\mathrm{ARMA}(r, r-1)$ model, so $\mathrm{GIC}(2r-1, 3)$.

D. Automatic ARMAsel Identification

The ARMAsel algorithm [28], [31] computes $\mathrm{AR}(p)$ models with $p = 0, 1, \ldots, N/2$ and selects a single best AR model with CIC (4). Generally, the highest candidate order p is limited to 1000 for $N > 2000$, but that is not necessary. Also $\mathrm{MA}(q)$ models are calculated with $q = 1, \ldots, N/5$, with an upper limit 400, and the best MA model is selected from those candidates with $\mathrm{GIC}(q, 3)$ of (9). The maximum MA order that is considered for computation is much lower than the maximum AR order, but practical data generally do not have very high MA orders. Moreover, a high AR order model is used as intermediate for the estimation of MA models. Also, $\mathrm{ARMA}(r, r-1)$ models are estimated for $r = 2, \ldots, N/10$ with the maximum 200 and the best ARMA order is selected, which has the smallest value for the criterion $\mathrm{GIC}(2r-1, 3)$. Finally, having selected the best $\mathrm{AR}(p)$ model, the best $\mathrm{MA}(q)$ model and the best $\mathrm{ARMA}(r, r-1)$ model, the prediction error of those three resulting models is estimated with the given data [3]. For MA and ARMA models this is given by

$$\mathrm{PE}(m) = \{\mathrm{RES}(m)\} \frac{1 + \frac{m}{N}}{1 - \frac{m}{N}} \qquad (15)$$

where m is the number of estimated parameters in the model. For $\mathrm{AR}(p)$ models, the expression is given by the finite sample expression [3]

$$\mathrm{PE}(p) = \{\mathrm{RES}(p)\} \prod_{m=1}^{p} \frac{1 + \frac{1}{N+1-m}}{1 - \frac{1}{N+1-m}} \qquad (16)$$

which differs only significantly from (15) if the number of estimated parameters is greater than $N/10$. The model type with the smallest estimate for the prediction error is selected. In this way, a single time series model, with selected type and order, can be determined for the given observations: it is called the ARMAsel model. The spectral density or the covariance can be computed from its estimated parameters, and that gives the best representation of the second-order characteristics of the measured data.

The new ARMAsel algorithm has many applications and is an example of intelligent signal processing. Its use does not require a profound knowledge of the user. All theoretical elements have been integrated in the Matlab toolbox ARMASA [31], with main routine ARMAsel. Of course, the experimenter can overrule many choices that are made automatically in the algorithm. It returns the parameters of the selected model, and the information of a number of additional models is also available from the algorithm upon request.

VII. Applications

With present-day computers, it is feasible to apply this automatic time series identification to measured data if an accurate spectral estimate is desired. It takes less time to compute tapered and windowed periodogram estimates, but those can never compete in accuracy with the unique automatically selected ARMAsel model. This has been proved theoretically, and it has been demonstrated in numerous simulations. As far as the author is aware, the algorithm of Section VI is the first successful attempt to automatically identify a time series model for measured observations without interaction from the user. The Matlab program ARMAsel needs only the observations as input, and it gives the AR and the MA parameters of the selected model as output. The algorithm is freely available [31].

A number of applications of the automatic ARMAsel algorithm have been studied already or are under study.

- Detection of methacholine from lung sounds[32];
- detection of objects in radar clutter[33];
- detection of the flow regime in turbulent flow;
- spectral representation of irregularly sampled data [34];
- spectral estimation with missing data;
- spectral representation of meteorological data;
- improved estimate for the accuracy of the estimated mean value, with this accuracy computed from the sum of estimated covariances [35];
- comparison of different spectral models with the single number ME (model error (4)) that is calibrated;
- extracting the statistically significant peaks from data;
- giving a brief description of the statistically significant details, which diminishes the number of bits that have to be transmitted in a remote sensing;
- improving the accuracy of most applications that select time series models with neural networks, by using the ARMAsel model that has statistically optimal properties and does not require learning sets with all kinds of problems about which sets to use for learning;
- monitoring the rotating and vibrating equipment;
- in general, monitoring to detect statistically significant but unmodeled and unexpected changes in the process behavior in a stochastic environment.

The example in Fig. 1 shows the automatically selected estimated ARMAsel spectra for $N = 100$ and for $N = 1000$ of some AR(21,20) process. Both spectra are rather close to the best achievable accuracy for that N, given by $\mathrm{ME}_{\min} = p+q = 41$. Clearly, more observations give better results. If initial or standard model parameters are available, newly selected ARMAsel spectra are an excellent process characterization to be used for monitoring and detection by looking for changes with the model error ME.

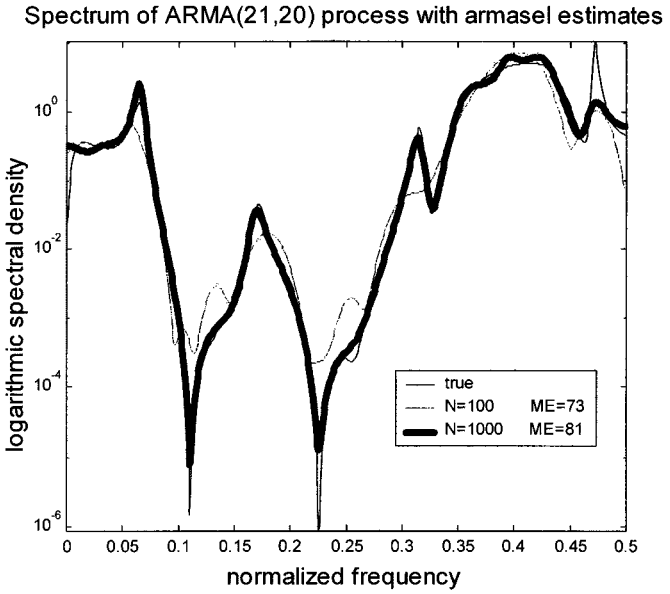

Fig. 1. True spectrum and two estimated ARMAsel spectra.

VIII. Concluding Remarks

It is feasible to estimate $N/2$ AR models, $N/5$ MA models and, $N/10$ ARMA models from N measured observations and to select automatically a single good model. To reduce the computation time, it may be advisable to restrict those upper boundaries for N greater than 500 or 1000. First, order selection is applied separately to the three model types AR, MA, and ARMA. The prediction errors of the selected $\mathrm{AR}(p)$, $\mathrm{MA}(q)$, and $\mathrm{ARMA}(r, r-1)$ models are used for selection of the model type with the ARMAsel algorithm.

Use of robust algorithms is necessary for the automatic identification. Observed differences can be attributed only to statistical variations as a function of the model order. The accuracy of the model with selected type and order is slightly worse than the accuracy of the model that could be estimated if type and order would *a priori* be known. For small samples, the accuracy may even be better.

ARMAsel is freely available [31]. Hence, it is easy to verify or to falsify the qualities of the ARMAsel algorithm by comparing the single ARMAsel spectrum with the result of one's own favorite spectral estimator.

References

[1] M. B. Priestley, *Spectral Analysis and Time Series*. New York: Academic, 1981.
[2] B. Porat, *Digital Processing of Random Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[3] P. M. T. Broersen, "Facts and fiction in spectral analysis," *IEEE Trans. Instrum. Meas.*, vol. 49, pp. 766–772, Aug. 2000.
[4] ——, "Finite sample criteria for autoregressive order selection," *IEEE Trans. Signal Processing*, vol. 48, pp. 3550–3558, Dec. 2000.
[5] P. M. T. Broersen and S. de Waele, "Windowed periodograms and moving average models," in *Proc 39th CDC Conf.*, Sydney, Australia, 2000, pp. 2706–2709.
[6] P. M. T. Broersen, "Selection of type and order of time series models," in *Proc 38th CDC Conf.*, Phoenix, AZ, 1999, pp. 5064–5069.
[7] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Comput.*, vol. 19, pp. 297–301, 1965.
[8] A. Schuster, "On the investigation of hidden periodicities with application to a supposed twenty-six day period of meteorological phenomena," *Terrest. Magn.*, vol. 3, pp. 13–41, 1898.
[9] G. U. Yule, "On a method of investigating periodicitics in disturbed series with special reference to Wolfer's sunspot numbers," *Phil. Trans. R. Soc. A*, vol. 226, pp. 267–298, 1927.
[10] E. Slutsky, "The summation of random causes as the source of cyclical processes," *Econometrica*, vol. 5, pp. 105–146, 1937.
[11] N. Wiener, "Generalized harmonic analysis," *Acta Math.*, vol. 35, pp. 117–258, 1930.
[12] A. Khintchine, "Korrelationstheorie der stationären stochastischen Prozessen," *Math. Ann.*, vol. 109, pp. 604–615, 1934.
[13] H. B. Mann and A. Wald, "On the statistical treatment of linear stochastic difference equations," *Econometrica*, vol. 11, pp. 173–220, 1943.
[14] P. Whittle, "Estimation and information in stationary time series," *Ark. Matematik*, vol. 2, pp. 423–434, 1953.
[15] S. Zacks, *The Theory of Statistical Inference*. New York: Wiley, 1971.
[16] M. Arato, "On the sufficient statistics for stationary Gaussian random processes," *Theory Probabil. Applicat.*, vol. 6, pp. 199–201, 1961.
[17] J. Durbin, "Efficient estimation of parameters in moving average models," *Biometrika*, vol. 46, pp. 306–316, 1959.
[18] ——, "The fitting of time series models," *Rev. Inst. Int. Stat.*, vol. 28, pp. 233–243, 1960.
[19] J. P. Burg, "Maximum entropy spectral analysis," in *Proc. 37th Meeting Soc. Exploration Geophys.*, Oklahoma City, OK, 1967, pp. 1–6.
[20] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control* San Francisco, CA, 1976.
[21] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716–723, Dec. 1974.
[22] E. Parzen, "Some recent advances in time series modeling," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 723–730, Dec. 1974.
[23] S. M. Kay and S. L. Marple, "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, pp. 1380–1419, 1981.
[24] P. M. T. Broersen, "The quality of models for ARMA processes," *IEEE Trans. Signal Processing*, vol. 46, pp. 1749–1752, June 1998.
[25] P. M. T. Broersen and S. de Waele, "Empirical time series analysis and maximum likelihood estimation," in *Proc. 2nd IEEE Benelux Signal Proc. Symp. SPS-2000*, 2000, pp. 1–4.
[26] J. S. Erkelens and P. M. T. Broersen, "Bias propagation in the autocorrelation method of linear prediction," *IEEE Trans. Speech Audio Processing*, vol. 5, pp. 116–119, Mar. 1997.
[27] P. M. T. Broersen, "Autoregressive model orders for Durbin's MA and ARMA estimators," *IEEE Trans. Signal Processing*, vol. 48, pp. 2454–2457, Aug. 2000.
[28] ——, "Robust algorithms for time series models," in *Proc. ProRISC/IEEE CSSP98*, 1998, pp. 75–82.
[29] P. M. T. Broersen and H. E. Wensink, "On the penalty factor for autoregressive order selection in finite samples," *IEEE Trans. Signal Processing*, vol. 44, pp. 748–752, Mar. 1996.
[30] ——, "Autoregressive model order selection by a finite sample estimator for the Kullback-Leibler discrepancy," *IEEE Trans. Signal Processing*, vol. 46, pp. 2058–2061, July 1998.
[31] P. M. T. Broersen. Matlab Toolbox ARMASA. [Online]. Available: http://www.tn.tudelft.nl/mmr/
[32] P. M. T. Broersen and S. de Waele, "Detection of methacholine with time series models of lung sounds," *IEEE Trans. Instrum. Meas.*, vol. 49, pp. 517–523, June 2000.
[33] S. de Waele and P. M. T. Broersen, "Modeling radar data with time series models," in *Eusipco 2000, Proc. 10th Eur. Signal Proc. Conf.*, Tampere, Finland, 2000, pp. 1–4.
[34] ——, "Error measures for resampled irregular data," *IEEE Trans. Instrum. Meas.*, vol. 49, pp. 216–222, Apr. 2000.
[35] P. M. T. Broersen, "Estimation of the accuracy of mean and variance of correlated data," *IEEE Trans. Instrum. Meas.*, vol. 47, pp. 1085–1091, Oct. 1998.

Piet M.T. Broersen was born in Zijdewind, The Netherlands, in 1944. He received the M.Sc. degree in applied physics in 1968 and the Ph.D. degree in 1976, both from the Delft University of Technology, Delft, the Netherlands.

He is currently with the Department of Applied Physics, Delft University of Technology. His main research interest is automatic identification on statistical grounds. He found a solution for the selection of order and type of time series models and the application to spectral analysis, model building, and feature extraction. His next subject is the automatic identification of input–output relations with statistical criteria.