TABLE I
STANDARD DEVIATIONS

Method	Bearing, Degrees		
	60°	80°	100°
SVR/ RANDOM	.2862	.2092	.2506
SVR/ DETERMINISTIC	.3129	.2451	.2846
FREQUENCY	5.590	4.578	6.107
SPAT. SMOOTH.	.3536	.3306	.3082

In the next example, the signal vector randomization method (SVR1) as given in (9), the deterministic SVR method (SVR2) as given in (11), the frequency estimation method (TK) outlined in Section III, and the spatial smoothing methods are compared.

All three sources are perfectly correlated, i.e., they have a zero relative phase angle at each iteration, e.g., multipath propagation where the path delays are the same. The analyses using SVR1, SVR2, and TK start by estimating the bearings with the frequency estimation method using one snapshot. The analysis using spatial smothing uses the first ten snapshots to estimate the spatial smoothing (SS) matrix. The results in Fig. 2 indicate the following.

- The frequency method is less accurate than the SVR and SS methods at an SNR of 10 dB. This is because the TK method uses a single snapshot, whereas SVR uses a weighted sum of, effectively, $[1/(1-\epsilon)]$ snapshots ($[\nu]$ is the smallest integer larger than ν), which is the equivalent window width for exponential averaging.
- The performance of the SVR1, SVR2, and SS methods is about the same.

In the final example, bearings are tracked for 500 snapshots. The SNR is 10 dB, and the sources are stationary at 60°, 80°, and 100°. In Table I, standard deviations for each of the three source bearings are listed for the SVR1, SVR2, frequency estimation, and spatial smoothing methods.

V. CONCLUSION

Two algorithms for the tracking of correlated signals have been presented. A frequency estimator is used to obtain the initial estimate of the DOA's in the presence of signal correlation. Signal vector randomization (SVR) can be employed to eliminate any correlation during the subspace tracking process, and this can be done automatically without any knowledge of the presence or absence of correlation. The SVR method does not reduce the effective array aperture as the spatial smoothing method does, and this becomes important in situations where the number of sensors in the array is small.

REFERENCES

[1] J. Cadzow, "Multiple source location—The signal subspace approach," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1110–1125, July 1990.
[2] S. Pillai and B. Kwon, "Performance analysis of MUSIC-type high resolution estimators for direction finding in correlated and coherent scenes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1176–1189, Aug. 1989.
[3] B. Friedlander and A. Weiss, "Direction finding using spatial smoothing with interpolated arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, pp. 574–587, Apr. 1992.
[4] T. Shan, M. Wax, and T. Kailath, "On Spatial smoothing for directions of arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 806–811, Apr. 1985.
[5] T. Shan and T. Kailath, "Adaptive beamforming for correlated signals and interference," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 527–536, June 1985.
[6] S. Valaee and P. Kabal, "The optimal focusing subspace for coherent signal subspace processing," *IEEE Trans. Signal Processing*, vol. 44, pp. 752–756, Mar. 1996.
[7] J. Yang and M. Kaveh, "Coherent signal subspace transformation beamformer," *Proc. IEEE*, vol. 78, pp. 267–275, Aug. 1990.
[8] M. Lu and Z. He, "Adaptive beamforming using split-polarity transformation for correlated signal and interference," *IEEE Trans. Antennas Propagat.*, vol. 41, pp. 314–324, Mar. 1993.
[9] F. Qian and B. V. Veen, "Partially adaptive beamforming for correlated interference rejection," *IEEE Trans. Signal Processing*, vol. 43, pp. 506–515, Feb. 1995.
[10] G. Golub and C. V. Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1983.
[11] D. Tufts, A. Kot, and R. Vaccaro, "The threshold effect in signal processing algorithms which use an estimated subspace," in *SVD and Signal Processing II: Algorithms Analysis and Applications*, R. Vaccaro, Ed. New York: Elsevier, 1991.
[12] D. Tufts and C. Melissinos, "Threshold extension of SVD-based algorithms," in *Proc. ICASSP*, pp. 2825–2828, 1988.
[13] D. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, pp. 975–989, Sept. 1982.
[14] C. MacInnes and R. Vaccaro, "Tracking directions-of-arrival with invariant subspace updating," *Signal Process.*, vol. 49, 1996.

The Quality of Models for ARMA Processes

P. M. T. Broersen

***Abstract*—The model error (ME) is an objective measure for assessing the quality of different models of a given ARMA process. The expression for ME can be evaluated easily in the time domain. This quality measure for known and given processes is necessary for an objective comparison of the performance of estimation algorithms and of order selection criteria.**

I. INTRODUCTION

A measure for the quality of models is important in a comparative study of algorithms in statistical time series. It is necessary for an objective comparison of the performance of the various estimation methods and of order selection criteria. The estimation algorithms and order selection criteria that provide models with the best average quality over a range of known processes are expected to also be the best when applied to unknown data. Classic measures are a tradeoff of bias and variance, but for specific purposes like control, a variety of error quantifications has been developed [1].

A measure for the accuracy of models for ARMA processes is introduced: the model error ME. The model error is based on the

Manuscript received June 9, 1997; revised November 11, 1997. The associate editor coordinating the review of this paper and approving it for publication was Prof. James A. Bucklew.

The author is with the Department of Applied Physics, Delft University of Technology, Delft, The Netherlands.

Publisher Item Identifier S 1053-587X(98)03945-2.

prediction error, which is defined as the mathematical expectation of the squared fit of an estimated model to new data that are generated by the same process as the data used for estimation of the model. ME can be applied to ARMA models as well as to pure AR or MA models by using zero-valued parameters. Moreover, it can also be used for the evaluation of the periodogram by determining the unique MA representation of the finite correlation function involved [2]. Several computational elements for the new quality measure have been described before [3], [4]. The result was a computer program or a labor-intensive recursive scheme. This analysis gives an easy computation and interpretation of the prediction error in the time domain as the variance of a specific ARMA process. The Fourier transform theory shows that the minimum of the prediction error agrees in the frequency domain with the maximum whiteness or flatness of the spectrum of the prediction error. Therefore, the integrated ratio of the true and the estimated spectrum over the whole frequency interval is as close as possible to one for the model with the smallest prediction error or model error.

Finally, the results of simulations demonstrate the use of an objective measure. The average quality of estimated and selected AR, MA, and ARMA models is evaluated. In addition, the influence of the length of the windowed correlation function on periodogram estimates [5] is shown and compared with the popular approach with multiple prolate spheriodal tapers [6].

II. Accuracy of Prediction

In studying algorithms for statistical signal processing, it is important to have an objective measure for the quality of different models for a known process to calibrate estimation algorithms and selection criteria as well as to make mutual comparisons. That evaluation for known processes can be used as a guideline to decide what can be expected when algorithms and order selection criteria are applied to unknown data. In many applications, the model that can predict future values of the time series with the highest accuracy will be a good model. Those applications include

- prediction in the time domain;
- parametric spectral modeling;
- the detection of errors or changes;
- the classification of data into categories;
- the quantization of parameters for speech coding.

For AR models from AR processes, a quality measure has been introduced that has been used to detect the differences between estimation methods in finite samples [7]. It has a sensible interpretation in the time and in the frequency domain, and it can be computed efficiently in the time domain. A similar measure will be defined for the more general class of ARMA processes.

Suppose that the ARMA(p, q) process xn is given by [5]:

$$\boldsymbol{A}(z)\mathrm{x}_{\boldsymbol{n}} = \boldsymbol{B}(z)\varepsilon_{\boldsymbol{n}} \tag{1}$$

or

$$\begin{aligned}\boldsymbol{x_n} + \boldsymbol{a_1 x_{n-1}} + \cdots + \boldsymbol{a_p x_{n-p}} \\ = \varepsilon_{\boldsymbol{n}} + \boldsymbol{b_1}\varepsilon_{\boldsymbol{n-1}} + \cdots \boldsymbol{b_q}\varepsilon_{\boldsymbol{n-q}}\end{aligned} \tag{2}$$

where ε_n is a white noise sequence with variance $\boldsymbol{\sigma}_{\varepsilon}^2$, and $A(z)$ is defined as [5]

$$\boldsymbol{A}(z) = \mathbf{1} + \boldsymbol{a_1} z^{-1} + \boldsymbol{a_2} z^{-2} + \cdots + \boldsymbol{a_p} z^{-\boldsymbol{p}}. \tag{3}$$

Likewise, a polynomial $B(z)$ is defined as a polynomial in z^{-1} of order q. The process is stationary if $A(z)$ has no poles for z on or outside the unit circle. Model parameters $A'(z)$ and $B'(z)$ of orders p' and q', respectively, which are not necessarily equal to p and q, are generally found by estimation of parameters from N observations $y_n, n = 1, \cdots, N$; the observations can be a realization of the process (1), but this is not necessary. Future predictions of the process (1) with this model can be made by substitution of those estimated parameters into new data x_n, which can be written as

$$\boldsymbol{B}'(z)\hat{\varepsilon}_n = \boldsymbol{A}'(z)\boldsymbol{x_n} \tag{4}$$

where the signal $\hat{\varepsilon}_n$ can be seen as the output of the estimated model with x_n used as input signal; the estimated $B'(z)$ is used as AR part and $A'(z)$ as MA part in this derivation of $\hat{\varepsilon}_n$. The squared error of prediction PE is defined as the expectation $E[\hat{\varepsilon}_n^2]$ with the condition that x_n is a realization of the process (1) that is independent of the observations y_n that had been used to estimate the parameters. The output $\hat{\varepsilon}_n$ of the model with $A'(z)$ and $B'(z)$ can be found by substituting x_n with (1)

$$\hat{\varepsilon}_n = \frac{\boldsymbol{A}'(z)}{\boldsymbol{B}'(z)}\,\boldsymbol{x_n} = \frac{\boldsymbol{A}'(z)\boldsymbol{B}(z)}{\boldsymbol{A}(z)\boldsymbol{B}'(z)}\,\varepsilon_{\boldsymbol{n}} = \frac{\boldsymbol{D}(z)}{\boldsymbol{C}(z)}\,\varepsilon_n. \tag{5}$$

Therefore, the relation between the error output $\hat{\varepsilon}_n$ of the model (4) and the innovations ε_n that generated the true process is given by an ARMA$(p+q', p'+q)$ process $D(z)/C(z)$. By considering (5) as a filtering operation [5], it follows elementarily that the covariance of $\hat{\varepsilon}_n$ is given as the inverse discrete-time Fourier transform of the power spectral density of a process with the white noise ε_n as input. This yields

$$\boldsymbol{R}_{\hat{\varepsilon}\hat{\varepsilon}}(\boldsymbol{\tau}) = \frac{\boldsymbol{\sigma}_{\hat{\varepsilon}}^2}{\mathbf{2\pi}} \int_{-\boldsymbol{\pi}}^{\boldsymbol{\pi}} \left| \frac{\boldsymbol{A}'(e^{\boldsymbol{j\omega}})\boldsymbol{B}(e^{\boldsymbol{j\omega}})}{\boldsymbol{A}(e^{\boldsymbol{j\omega}})\boldsymbol{B}'(e^{\boldsymbol{j\omega}})} \right|^2 e^{\boldsymbol{j\omega\tau}}\, \boldsymbol{d\omega} \tag{6}$$

if all poles of $A(z)B'(z)$ are inside the unit circle, which is assumed throughout this correspondence. In the derivation of (6), from (5), it is essential that the estimated polynomials $A'(z)$ and $B'(z)$ are stochastically independent of ε_n. Hence, (6) does not describe the covariance of the residuals that are minimized in the estimation of the parameters because the estimated parameters are then correlated with the specific realization of ε_n that was used for the estimation. However, the squared error of prediction or, in short, the prediction error PE, because it is the variance of $\hat{\varepsilon}_n$, follows from (6) for $\tau = 0$. That integral can be approximated by a finite sum in the frequency domain, which will be quite accurate unless the true poles or the estimated zeros are close to the unit circle. The integral can also be solved with calculus in the complex plane. An efficient numerical solution for a similar complex integral has been given [3] as has a time domain solution [4]. Those solutions are available as computer programs or as a complicated recursive schemes. An easier derivation of the same result in the time domain is given here with some filter theory.

The output of the ARMA process in (5) can be described with a realization of the filter operation by separating it in two consecutive filters with an intermediate signal v_n as

$$\begin{aligned}\boldsymbol{v_n} + \boldsymbol{c_1 v_{n-1}} + \cdots + \boldsymbol{c_{p+q'} v_{n-p-q'}} = \varepsilon_{\boldsymbol{n}} \\ \hat{\varepsilon}_{\boldsymbol{n}} = \boldsymbol{v_n} + \boldsymbol{d_1 v_{n-1}} + \cdots + \boldsymbol{d_{p'+q} v_{n-p'-q}}.\end{aligned} \tag{7}$$

The covariance $R_{vv}(k) = E[v_n v_{n+k}]$ can be computed with the well-known Yule–Walker relations for autoregressive processes [5]. Extra terms of $R_{vv}(k)$ for $k > p + q'$ can be computed by extrapolating the autoregressive covariance function with the Yule–Walker relations [5, p. 133]. Afterwards, the second relation in (7) can be used to express the covariance $E[\hat{\varepsilon}_n \hat{\varepsilon}_{\boldsymbol{n+k}}]$ in the covariances of v_n, yielding (8), shown at the bottom of the page. The squared error of prediction PE, which is defined as the expectation $E[\hat{\varepsilon}_{\boldsymbol{n}}^{\boldsymbol{2}}]$ or $\boldsymbol{\sigma}_{\hat{\varepsilon}}^{\boldsymbol{2}}$, follows as the result of this equation for $k = 0$. The interpretation of PE in the time domain and in the frequency domain is illustrated in (6) for $\tau = 0$. The model with the smallest possible value for $\boldsymbol{\sigma}_{\hat{\varepsilon}}^{\boldsymbol{2}}$ (a minimum that is equal to $\boldsymbol{\sigma}_{\varepsilon}^{\boldsymbol{2}}$) is also the model where the integrated ratio of the true and the estimated spectrum in the frequency domain equals one everywhere with an equal numerator and denominator polynomial [8].

III. Model Error

The asymptotical value for the prediction error PE is $\boldsymbol{\sigma}_{\varepsilon}^{\boldsymbol{2}}\ \{1 + (p' + q')/N\}$ for all models where $p' \geq p$ and $q' \geq q$ and if $A'(z)$ and $B'(z)$ of the model are estimated from N observations. Those complete or overcomplete models have no bias contribution to PE, and each estimated parameter adds $\boldsymbol{\sigma}_{\varepsilon}^{\boldsymbol{2}}/N$ due to its variance. The constant $\boldsymbol{\sigma}_{\varepsilon}^{\boldsymbol{2}}$ in the PE is given by the variance of the innovations in (1). This scaling factor and constant contribution $\boldsymbol{\sigma}_{\varepsilon}^{\boldsymbol{2}}$ gives no relevant information about the quality of different models for the same process. Moreover, the sample size N has influence on the PE. However, it is often important that estimation algorithms have a good performance for all sample sizes. To obtain a measure for the model quality that is not explicitly dependent on the sample size, at least not in the asymptotic theory, the ME is defined as

$$\boldsymbol{ME} = \boldsymbol{N}\left(\frac{\boldsymbol{PE} - \boldsymbol{\sigma}_{\varepsilon}^{\boldsymbol{2}}}{\boldsymbol{\sigma}_{\varepsilon}^{\boldsymbol{2}}}\right) = \boldsymbol{N}\left(\frac{\boldsymbol{\sigma}_{\hat{\varepsilon}}^{\boldsymbol{2}}}{\boldsymbol{\sigma}_{\varepsilon}^{\boldsymbol{2}}} - \boldsymbol{1}\right). \tag{9}$$

The ME is a scaled version of the prediction error after subtraction of the innovation variance. The asymptotical expectation is $p' + q'$ for models where $p' \geq p$ and $q' \geq q$. The ME will be positive if all poles of $C(z)$ are strictly inside the unit circle; otherwise, the ME gives no useful result. It can only become zero if the polynomials $C(z)$ and $D(z)$ in (5) are equal; otherwise, it can be shown that it is always positive [8]. The ME is a useful measure for the model accuracy, with a significance in time and in frequency domain. The ME simplifies to an autoregressive measure [7] when the MA parameters are absent, and it can also be used for the fit of the MA models to the MA data by using zero values for the AR parameters.

The ME has been used in the statistical evaluation of estimation algorithms and order selection criteria for MA models [9], for ARMA models [10], and for a comparison of AR and MA models. An important result is that the ME approaches ∞ if an estimated MA polynomial in $C(z)$ has zeros that approach the unit circle, unless of course they are canceled by poles in $D(z)$. This means that MA or ARMA algorithms that can possibly produce estimated zeros on or even outside the unit circle (without the process having a zero there) have an objective average quality ME that will become infinitely bad for some processes.

IV. Simulation Results

Simulations have been carried out with an ARMA$(2, 1)$ process given by

$$\boldsymbol{x_n + 0.39x_{n-1} + 0.30x_{n-2} = \varepsilon_n - 0.90\varepsilon_{n-1}}. \tag{10}$$

Different spectral models have been compared by determining the average ME of ARMA(2, 1) models and of estimated AR, MA, and ARMA models with selected orders. The models are

- AR model estimated with the Burg technique [2] and with order selected with penalty factor 3 in the order selection criterion [11];
- MA model estimated with a sliding window modification of Durbin's method and MA order selected with penalty factor 3 [9];
- ARMA(2, 1) model estimated with a sliding window improvement on Durbin's iterative method of sequentially updating AR and MA parts of the estimate [10];
- ARMA model estimated like the previous one but with order p selected with penalty factor 3 from estimated ARMA$(p, p-1)$ models;
- PER$(N/2)$: periodogram with a cosine taper over the first and last 10% of the data and with Parzen window [5] with length $N/2$ over the correlation function;
- similar periodograms with cosine taper and Parzen window with lengths $N/4, N/8,$ and $N/16$;
- PER(mtap): periodogram with multiple prolate spheriodal tapers [6]; the given periodogram is the mean of the periodograms with the first seven spheroidal tapers.

Each periodogram has been computed with at least as many zeros added to the signal as the original length of the observations to ensure that the inverse Fourier transform of the periodogram was the biased estimate of the true covariance function [5]. This covariance function, of length $N - 1$ or less if a window is used, has an unique relation with a MA model [2]. The parameters of that MA model can be determined with a nonlinear factorization algorithm [12]. This MA representation for given data can be used to compute the quality of the spectral periodogram estimate with the model error ME. In this way, periodograms and time series models are compared objectively.

The use of an objective measure is demonstrated in Fig. 1, where different spectral representations for one single realization of the ARMA process of (10) are given. The estimated variance of the process is the same for all methods; therefore, the variance can be left out of the comparison, and the normalized spectral densities are shown in Fig. 1. The ME values for this realization are 14.5, 3.6, 13.8, and 11.6 for AR, ARMA, PER$(N/8)$, and PER(mtap), respectively. The ME of the periodogram with only a taper was 430.1 in this example. Therefore, each of the four presented spectra is a much better approximation than the periodogram. However, a mutual comparison of the spectra based on the lines in Fig. 1 is quite difficult: A certain ratio between true and estimated spectra looks

$$\boldsymbol{R_{\varepsilon\hat{\varepsilon}}(k)} = (\boldsymbol{1}\ \boldsymbol{d_1} \cdots \boldsymbol{d_{p'+q}}) \begin{pmatrix} \boldsymbol{R_{vv}(k)} & \boldsymbol{R_{vv}(k+1)} & \cdots & \boldsymbol{R_{vv}(k+p'+q)} \\ \boldsymbol{R_{vv}(k-1)} & \boldsymbol{R_{vv}(k)} & & \\ \vdots & \vdots & \ddots & \\ \boldsymbol{R_{vv}(k-p'-q)} & \boldsymbol{R_{vv}(k-p'-q-1)} & \cdots & \boldsymbol{R_{vv}(k)} \end{pmatrix} \begin{pmatrix} \boldsymbol{1} \\ \boldsymbol{d_1} \\ \vdots \\ \boldsymbol{d_{p'+q}} \end{pmatrix} \tag{8}$$

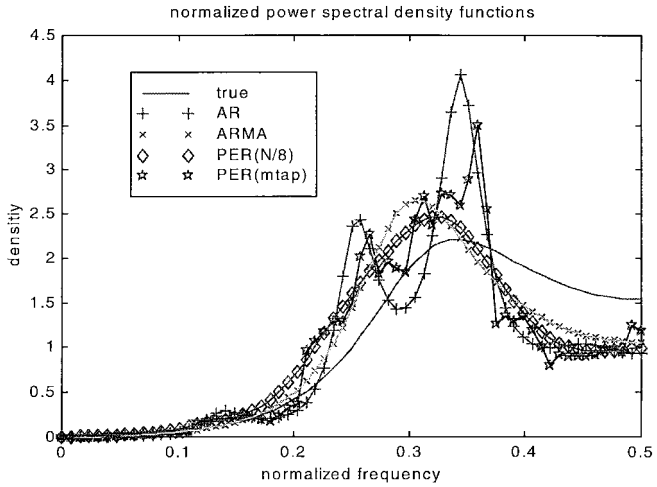

Fig. 1. True power spectral density of an ARMA(2, 1) process and four estimates obtained from one realization of 128 observations.

TABLE I
AVERAGE ME OF NINE SPECTRAL ESTIMATES AS A FUNCTION OF THE SAMPLE SIZE N IN 5000 SIMULATION RUNS WITH AN ARMA(2, 1) PROCESS

N	32	64	128	256	512
AR	13.2	15.3	18.1	21.2	23.9
MA	9.3	9.2	8.9	8.7	7.4
ARMA(2,1)	6.4	5.1	4.2	4.1	3.5
ARMA	8.6	6.2	4.8	4.5	3.7
PER(N/2)	7.7	11.3	21.0	41.3	84.6
PER(N/4)	9.7	8.3	11.1	19.9	40.1
PER(N/8)	23.7	15.3	10.8	11.3	19.2
PER(N/16)	46.8	45.8	27.5	16.3	12.8
PER(mtap)	8.8	8.7	12.5	21.9	43.0

much smaller for normalized frequency 0.1 than for 0.3, but it has the same influence on the prediction error, as measured in the ME. It is already difficult to obtain a proper comparison of the quality of the different algorithms for a single realization, e.g., it seems surprising that the twisting multitaper approximation PER(mtap) is better than PER($N/8$), which is the periodogram with window $N/8$. It will almost be impossible to carry out a visual comparison for many realizations and for different sample sizes.

The average of ME for different spectral estimates in simulations is given in Table I. This objective quality measure for spectral models gives a good impression of the accuracy and of the influence of the sample size. By including N in the definition of ME in (9), the numbers in the table are all of the same order of magnitude and are easier to interpret than the prediction error PE, which depends much more strongly on the sample size. It is remarkable that the ME increases with the sample size N for AR models, remains more or less a constant for MA models, and decreases for ARMA models. The average difference between the ME of the true order ARMA(2, 1) model and of ARMA models with selected orders becomes smaller for greater N. The result of multitaper approximations is not very different from usual windowed approximations; PER(mtap) with the average of seven tapers remains close to PER($N/4$) for all sample sizes. The best length for windows depends on the sample size. Its optimum is about 16 for the first three columns in the table and something longer for more observations. It is neither a constant nor a fixed proportion of N; therefore, it would be impossible to determine the optimum for unknown data. The best choice for the window length gives a reasonable spectral estimate for all sample sizes for the process in (10), although the quality of the best choice decreases as a function of the number of observations. Estimated ARMA spectra are best for this true ARMA process. The quality of AR spectra is something less than that of the best window choice in periodograms, but it is better than that of the wrong choices.

However, a major disadvantage of the spectral descriptions based on the periodogram is the absence of a good and objective statistical rule for the choice of the window size for unknown data. Order selection criteria can be used in time series models to find good AR models [11], good MA models [9], and good ARMA models [10]. For the window size, there is no statistically reliable equivalent.

V. CONCLUSIONS

The model error is a useful measure for the quality of selected models for known ARMA processes. It is a normalized and scaled version of the prediction error, and it can be computed with an exact expression in the time domain. In the frequency domain, it represents a number that can be used for an objective comparison of the quality of spectral estimates from time series models and from windowed periodograms over the whole frequency interval.

REFERENCES

[1] B. Ninness and G. C. Goodwin, "Estimation of model quality," *Automatica*, vol. 31, pp. 1771–1797, 1995.
[2] S. M. Kay and S. L. Marple, "Spectrum analysis, a modern perspective," *Proc. IEEE*, vol. 69, pp. 1380–1419, 1981.
[3] K. J. Åström, E. I. Jury, and R. G. Agniel, "A numerical method for the evaluation of complex integrals," *IEEE Trans. Automat. Contr.*, vol. AC-15, pp. 468–471, 1970.
[4] S. Kay, "Generation of the autocorrelation sequence of an ARMA process," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 733–734, 1985.
[5] M. B. Priestley, *Spectral Analysis and Time Series*. New York: Academic, 1981.
[6] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. Cambridge, U.K.: Cambridge Univ. Press, 1993.
[7] P. M. T. Broersen and H. E. Wensink, "On finite sample theory for autoregressive model order selection," *IEEE Trans. Signal Processing*, vol. 41, pp. 194–204, 1993.
[8] K. J. Åström and T. Söderström, "Uniqueness of the maximum likelihood estimates of the parameters of an ARMA model," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 769–773, 1974.
[9] P. M. T. Broersen, "The best order of long autoregressive models for moving average estimation," in *Proc. Signal Processing VIII, Eusipco Conf.*, Trieste, Italy, 1996, pp. 799–802.
[10] ——, "On orders of long AR models for ARMA estimation," in *Proc. ECSAP Conf.*, Prague, Czech Republic, June 24–27, 1997, pp. 83–86.
[11] 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, 1996.
[12] G. Wilson, "Factorization of the covariance generating function of a pure moving average process," *SIAM J. Numer. Anal.*, vol. 6, pp. 1–7, 1969.