

The Performance of Spectral Quality Measures

Piet M. T. Broersen

Abstract—Two different classes of quality measures are discussed and compared: absolute and relative measures. The relative class, to which the prediction error belongs, has many different equivalent members, like the spectral distortion and the likelihood ratio. This type of measure is based on time series theory. The prediction error can be written either as a squared error of prediction in the time domain or as a relative error in the frequency domain. It is useful in many applications, especially in comparing models obtained with different estimation algorithms. It will be compared to some measures that are absolute in the frequency domain. To that class belongs the integrated squared difference between spectra, that gives equal weights to all frequencies. Another measure is based on the squared difference between impulse responses. The absolute class has only a few practical applications, mainly in speech.

Index Terms—Likelihood ratio, model accuracy, model error, prediction error, spectral distortion.

I. INTRODUCTION

AN OBJECTIVE measure for the quality of models is necessary for the mutual evaluation of estimation algorithms. Spectral quality measures have always been an important issue in speech processing. The properties of several related measures have been investigated theoretically and experimentally [1]. The treatment in the speech literature is generally limited to autoregressive (AR) processes or models, because the linear prediction models in speech processing are AR models. A close relation exists between several measures in the relative class: the prediction error, the likelihood ratio, the spectral distortion, and the cepstrum. They are all identical for practical purposes if the spectral distortion is less than 2 dB [1]. A similar measure can also be expressed in other representations of AR models, so in reflection coefficients, line spectrum pairs or log area ratios [2].

For combined autoregressive-moving average (ARMA) processes, the concept of the prediction error has been used to derive the *model error* [3]. If the true process parameters are known, as in simulations, the prediction error can be expressed as the variance of an ARMA process, whose parameters are cross convolutions of the true process polynomials with the estimated AR and MA polynomials [3]. For unknown process parameters, however, the prediction error remains a useful quantity with a definite practical meaning. It is the expectation of the squared one step ahead prediction with a time series model, under the condition that the predicted observation has not been used to estimate that model. A further advantage of the prediction error as a quality measure is that its relevance can be indicated both in the time and in the frequency domain. For AR

processes, it has been shown that the minimum prediction error variance is obtained if the spectrum of the residuals is maximally flat [4]. This, in turn, means that all frequencies are equally important, and the model with minimum prediction error also minimizes the *ratio* of the true and estimated spectral densities. This background of time and frequency domain is also clear in the derivation of the model error for ARMA processes [3]. The relationship between the prediction error and the integrated ratio of squared spectral densities has also been derived [5], [6]. That measure has recently been used for a comparison of the statistical accuracy of nonparametric spectral estimators [7]. This integrated squared ratio of spectra, with or without logarithm, also appears in an expression for the Cramér-Rao lower bound in parametric spectral estimation [8]. Some other relative precision measures with ratios of spectra have been proposed as percentage errors, signal-to-noise ratio (SNR) or degrees of freedom (DOF) [9].

Mathematically, *distance* measures must have the property of symmetry: the distance from A to B should be equal to the distance from B to A. It turns out that the relative measures prediction errors or likelihood ratios do not obey this requirement for distances because they are defined by quotients; spectral distortion is an example of a relative distance measure because it is based on a *squared* logarithmic difference.

Absolute measures are also distances. An absolute measure that is often mentioned in practice is the sum of squared differences between the true and the estimated covariance or correlation functions. With Parseval's relation, this is equal to the integrated mean squared difference between the true and the estimated spectrum. No claims for optimality from a theoretical point of view are known for this measure. It seems to be an obvious choice where the accent is on estimating the strong parts of the spectrum. However, the consequences are not always desirable because errors in weak parts of the spectrum have little influence on this distance measure for spectral quality. This measure is proposed as a figure of merit in estimating parameters from laser Doppler measurements [10], [11].

A completely different concept for absolute measures has been derived as the reconstruction error distortion (RED) [12]. The RED is defined as the sum of the squared differences between estimated and true impulse responses. Parametric impulse responses are computed by using the AR and MA polynomials with a single pulse as input signal. The RED is especially useful for vowels in speech processing. An explanation is that the model for vowels is in fact an impulse response to a sequence of pitches rather than an autoregressive response to a white noise excitation signal. The result may be that use of RED in codebook searching requires fewer bits, giving a transparent quantization for speech with a smaller size of the codebook [12]. The favorable performance of RED in speech coding is most likely due

Manuscript received May 26, 1999; revised February 28, 2001.

The author is with the Department of Applied Physics, Delft University of Technology, 2600 GA Delft, The Netherlands.

Publisher Item Identifier S 0018-9456(01)04392-3.

to the masking properties of the human ear, where soft parts in the spectrum can be masked by adjacent strong parts. This human sensitivity may favor absolute differences above relative in speech.

This paper gives an empirical comparison of the different types of spectral quality measures, especially of relative and absolute measures. An example shows that a model can have a very small absolute error, while the relative error is large. This result can be found in examples with a high dynamic range where there is a large difference between the power in weak and in strong parts of the spectrum.

II. QUALITY MEASURES

A. Absolute Measures

A class of absolute quality measures that can be seen as distances between two processes or between a process and its model is derived from Parseval's relation applied to covariances $R(i)$ and spectral densities $h(\omega)$, which form a Fourier transform pair. That relation applied to the difference between the true covariance function and an estimate that is indicated by $\hat{\cdot}$ gives

$$\sum_{i=-\infty}^{i=\infty} (\mathbf{R}(i) - \hat{\mathbf{R}}(i))^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} (h(\omega) - \hat{h}(\omega))^2 d\omega. \quad (1)$$

In practice, finite sums of covariances are used for computation of the left-hand side; the integral of the right-hand side is replaced by a finite summation obtained with the discrete Fourier transform. Generally, measures calculated by a summation in the frequency domain become inaccurate if poles of the discrete-time process or model are close to the unit circle because the continuous spectrum is not approximately constant over the elementary frequency interval of the discrete transform. In the time domain, a finite order computation with a summation of K covariances can only be a good approximation for (1) if the true covariance $R(i)$ is negligible for $i > K$. Even then, $\hat{R}(i)$ will never die out because it has a finite variance for $i > K$ [9].

The estimated variance of the signal will often be the same for different spectral estimators. Hence, it might be better to remove the variance from the spectral quality measure (1) by taking normalized correlations $\rho(i)$, with $\rho(0) = 1$. The normalized spectral densities $f(\omega)$ have the integral of the power equal to unity and give

$$\begin{aligned} \mathbf{IMSE} &= \sum_{i=-\infty}^{i=\infty} (\rho(i) - \hat{\rho}(i))^2 \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (f(\omega) - \hat{f}(\omega))^2 d\omega. \end{aligned} \quad (2)$$

IMSE is an acronym for integrated mean square error. The infinite sum of $R(i)^2$ has a practical meaning because it determines the variance of the estimate for $R(0)$ [9]. However, it is generally not possible to find a process that has the difference $R(i) - \hat{R}(i)$ in (1) as covariance or equivalently the difference between two spectra as spectral density; the difference between spectra can become negative and not be a spectrum itself. Therefore (1) and

(2) cannot be transformed into properties of processes. Nevertheless, the properties of $\hat{R}(i)$ as a function of i are important in that a valid covariance estimate should be positive semi-definite [9]. There are no constraints on the function in (1) and (2). As a consequence, estimated correlation functions that are not positive semi-definite can have a small IMSE, but they are useless because they can produce negative spectral densities for some frequencies. Furthermore, the most accurate estimator for the infinite sum of $\rho(i)^2$ has been studied in optimizing the accuracy of the estimated variance [13]. The best estimator for this infinite sum is *not* obtained from a summation of estimated correlations; it is difficult to decide how many terms should be included. It is better calculated with the parameters of a single time series model [13] estimated from the data with special estimation and order selection algorithms [14].

No objective description of the properties of the measure (1) is known, except that it seems to be an obvious choice in giving priority to estimating the strong parts of the spectrum. Recently, (2) has been used as a criterion for the analytical approximation of measured turbulence correlation functions by prototypes of the desired shape [10], [11]. Using $\{\rho(i) - \hat{\rho}(i)\}^2$ as criterion for minimization is not optimal in least squares theory because the errors in the estimated correlations are not independent, but they are strongly covariant [9]. An optimal solution would require generalized least squares with the inverse of the covariance matrix of the measurement errors as weighting matrix [15]. However, a transformation to independent measurement errors can also give an optimal solution. Modeling in the frequency domain would not have the problem of dependence between the measurement errors.

A second absolute quality measure is RED [12], which is special for speech coding. It uses a scaled sum of the squared differences between the true impulse response $g(i)$ and its estimate

$$\mathbf{RED} = \frac{\sigma_\varepsilon^2}{\sigma_x^2} \sum_{i=0}^{\infty} (g(i) - \hat{g}(i))^2. \quad (3)$$

The impulse responses can be computed with the AR and MA polynomials that will be defined in the next section using a pulse as input signal. With Parseval's relation and with (7), this expression can be written in the frequency domain as

$$\mathbf{RED} = \frac{\sigma_\varepsilon^2}{2\pi\sigma_x^2} \int_{-\pi}^{\pi} \left| \frac{\mathbf{B}(e^{j\omega})}{\mathbf{A}(e^{j\omega})} - \frac{\hat{\mathbf{B}}(e^{j\omega})}{\hat{\mathbf{A}}(e^{j\omega})} \right|^2 d\omega. \quad (4)$$

For AR processes, an easy computation of RED has been given [12]. The extension to ARMA processes is obvious with (4). Application to code book search in speech coding has shown that the use of either RED or relative measures gives a different selection of code words for the same speech signal. In the frequency domain, RED uses the absolute value of the difference in transfer functions. In that respect RED is related to the IMSE of (2). That also follows from the observation that the correlation is the convolution of $g(i)$ with $g(-i)$.

B. Relative Measures

The class of measures that are relative in the frequency domain contains the prediction error, the likelihood ratio and the

integrated ratio of squared spectral densities; those and other measures are asymptotically equivalent [1], [2]. Suppose that the ARMA(p, q) process x_n is given by [9]

$$\mathbf{A}(z)\mathbf{x}_n = \mathbf{B}(z)\varepsilon_n \quad (5)$$

or

$$\mathbf{x}_n + \mathbf{a}_1\mathbf{x}_{n-1} + \cdots + \mathbf{a}_p\mathbf{x}_{n-p} = \varepsilon_n + \mathbf{b}_1\varepsilon_{n-1} + \cdots + \mathbf{b}_q\varepsilon_{n-q}$$

where ε_n is an i.i.d. white noise sequence with variance σ_ε^2 and $A(z)$ is defined as

$$\mathbf{A}(z) = \mathbf{1} + \mathbf{a}_1z^{-1} + \mathbf{a}_2z^{-2} + \cdots + \mathbf{a}_pz^{-p}. \quad (6)$$

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. It is invertible if the roots of $B(z)$, denoted zeros, are all inside the unit circle. The spectral density of an ARMA process is given by [9]

$$\mathbf{h}(\omega) = \sigma_\varepsilon^2 \mathbf{f}(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\mathbf{B}(e^{-j\omega})|^2}{|\mathbf{A}(e^{-j\omega})|^2}. \quad (7)$$

The prediction error $\hat{\varepsilon}_n$ of a model of a given process is defined as the squared error when applying the estimated model to a new, statistically independent, realization of that process. A measured realization can be filtered with model parameters $\hat{\mathbf{A}}(z)$ and $\hat{\mathbf{B}}(z)$ of orders p' and q' , respectively, not necessarily equal to p and q . Predictions of the process (5) can be made by substitution of these model parameters into the measured x_n , which can be written as

$$\hat{\mathbf{B}}(z)\hat{\varepsilon}_n = \hat{\mathbf{A}}(z)\mathbf{x}_n \quad (8)$$

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 $\hat{\mathbf{B}}(z)$ is used as the AR part, and $\hat{\mathbf{A}}(z)$ as the MA part in this description 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 (5) that is independent of the observations that had been used to estimate the parameters. The output $\hat{\varepsilon}_n$ of the model (8) can be found by substituting x_n with the relation in (5)

$$\hat{\varepsilon}_n = \frac{\hat{\mathbf{A}}(z)}{\hat{\mathbf{B}}(z)}\mathbf{x}_n = \frac{\hat{\mathbf{A}}(z)\mathbf{B}(z)}{\mathbf{A}(z)\hat{\mathbf{B}}(z)}\varepsilon_n. \quad (9)$$

The relationship between the prediction $\hat{\varepsilon}_n$ and the innovations ε_n that generated the true process is therefore given by an ARMA($p + q', p' + q$) process. It follows [3] that the variance of ε_n is given as

$$\text{PE} = \sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{2\pi} \int_{-\pi}^{\pi} \left| \frac{\hat{\mathbf{A}}(e^{j\omega})\mathbf{B}(e^{j\omega})}{\mathbf{A}(e^{j\omega})\hat{\mathbf{B}}(e^{j\omega})} \right|^2 d\omega \quad (10)$$

if all poles of $A(z)\hat{\mathbf{B}}(z)$ are inside the unit circle, which is assumed throughout this paper. The constant scaling factor σ_ε^2 gives no relevant information about the quality of different models for the same process. Therefore, the ME for a model

estimated from N observations is made independent of σ_ε^2 as [3]

$$\text{ME} = \mathbf{N}(\text{PE}/\sigma_\varepsilon^2 - 1). \quad (11)$$

The asymptotic expectation of ME equals $p' + q'$ for a model that has all true nonzero parameters included, so with $p' \geq p$ and $q' \geq q$, independent of N , σ_ε^2 and the true process parameters [3]. This property of ME has advantages in the interpretation of simulation results. Above the true order, this quality measure is not dependent on the shape of the true spectrum. The ME should not be used if the estimated variance itself is important for the quality of the estimated spectrum.

The class of relative measures related to the PE and ME has many different members. One of them, the spectral distortion (SD) is defined as

$$\text{SD} = \frac{0.5}{2\pi} \int_{-\pi}^{\pi} \left[\ln\{\mathbf{f}(\omega)\} - \ln\{\hat{\mathbf{f}}(\omega)\} \right]^2 d\omega. \quad (12)$$

The SD can be expressed in the true and the estimated polynomials with (7). The SD can with (12) be computed over a part of the frequency domain. Equally important is that a variant of SD can be defined with $h(\omega)$ instead of $f(\omega)$ in (12). The variant with $h(\omega)$ is appropriate when the estimated variance is also a part of the comparison. This occurs in the comparison of interpolation schemes. Some simple interpolation methods maintain the true signal variance; and complex interpolation schemes, which use more points per interpolation, disturb that variance [16]. Interpolation is necessary for the use of fast Fourier transform (FFT) or time series analysis on irregularly sampled data, as in laser anemometry of turbulence. Another member of this group is the integrated *ratio* IRSE of squared spectral densities [5], [6] that is defined as

$$\text{IRSE} = \frac{0.5}{2\pi} \int_{-\pi}^{\pi} \left[\frac{\mathbf{f}(\omega) - \hat{\mathbf{f}}(\omega)}{\mathbf{f}(\omega)} \right]^2 d\omega. \quad (13)$$

Absolute and relative measures will be compared in the next section. Their performance is similar for processes with rather flat spectra, and the difference between the absolute measure (2) and the relative measure (13) is small. In white noise, $f(\omega)$ is a constant and (2) and (13) are, except for a scaling factor, identical. However, the difference becomes pronounced for spectra with high and sharp peaks and deep valleys. A turbulence spectrum will be treated as an example having a steep slope at high frequencies.

C. Rationale for Time Series Measures

At first sight, IMSE might look more general than ME and RED because IMSE is the integral of a spectrum, and ME and RED use time series models. A spectrum can be found from equidistant observations of arbitrary signals as the periodogram: the product of the FFT with its complex conjugate. The inverse transform of the periodogram is equal to the estimated covariance function if sufficient zeros have been added to the signal before applying the FFT algorithm [9]. However, any positive definite covariance function of finite length K can be interpreted as coming from an MA(K) model [4]. The K parameters can

ME, RED and IMSE of AR models for an AR(1) process, N=64

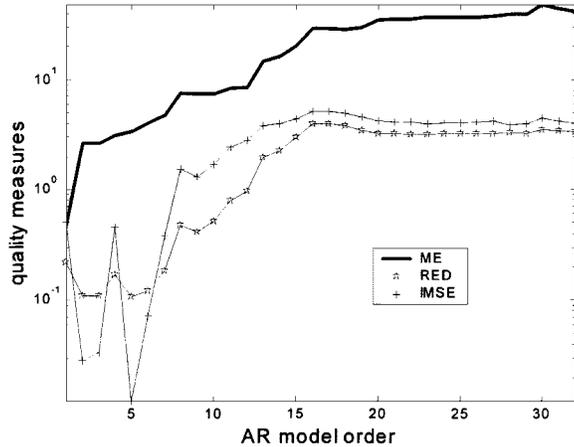


Fig. 1. Three types of quality measures as a function of the order of the AR model for 64 observations of an AR(1) process.

be found with an iterative algorithm [17]. Hence, a spectrum estimated from the periodogram of the data, with or without windowing and tapering, is equivalent to a finite order MA process. For that reason, time series models and measures are certainly not less general than measures derived from periodograms or spectra.

III. COMPARISON OF QUALITY MEASURES

The first comparison is made by evaluating the quality of estimated AR models as a function of the AR order. The models of increasing orders have been estimated from a realization of 64 observations of an AR(1) process with $a_1 = -0.9$. It is a priori obvious that the AR(1) model gives a good description for this process. The quality of the models will decrease for higher model orders. Fig. 1 shows three criteria: 1) ME, 2) RED multiplied by the constant $\sigma_x^2/\sigma_\varepsilon^2$, and 3) IMSE. The ME has its minimum at the expected order 1, RED at order 3, and IMSE at order 5. Moreover, ME is a continuously increasing function of the model order as expected. The RED and IMSE behave differently as a function of the AR model order, and they have several local minima. In this example, the properties of ME as a quality measure look more reliable. There is nothing in the residual variance or in the estimated reflection coefficients that can explain the irregularities of IMSE as a function of the AR model order of this AR(1) process. All frequencies contribute with the same weight to ME and equivalent measures. For IMSE, errors where $f(\omega)$ is high contribute much more than relative errors of the same magnitude for frequencies where $f(\omega)$ is small. Accidental and relatively small improvements of the fit in the strong parts of the spectrum explain this irregular behavior of the integral IMSE.

It is seen in Fig. 2 that ME, SD, and the relative squared spectral measure IRSE of (13) look similar and have their minimum for AR order 1, the true process order. This is in agreement

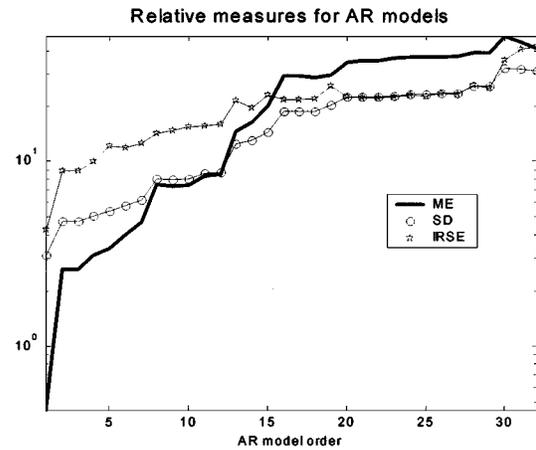


Fig. 2. Three relative quality measures as a function of the order of the AR model for 64 observations of an AR(1) process.

IMSE and its variant of Eq(1) for an AR(1) process, N=64

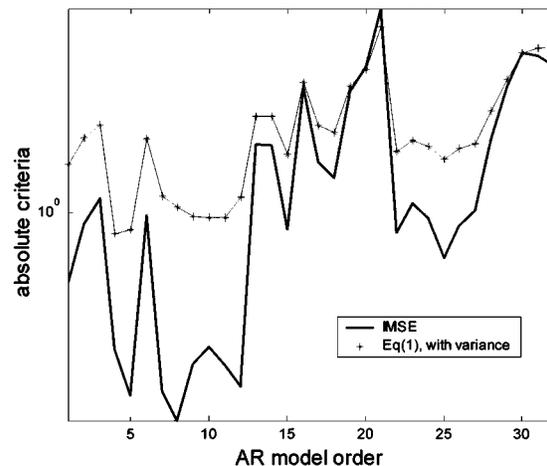


Fig. 3. Influence of using the estimated variance in the quality measures. The measures are a function of the order of the estimated AR model in a single simulation run. The true variance of the process is 5.26, the estimated variance 4.07.

with the asymptotical theory where those three measures become identical with the proper scaling constants.

Due to their definition, the estimated variance of the signal plays no role in ME and RED. Only the parameters of the ARMA process are required for those measures in this simulation study.

The error measure defined in (1) is similar to IMSE, but it also includes the measured variance. Fig. 3 shows that results of those criteria may be different. The IMSE has its minimum at order 8 in this simulation example, and the criterion of (1) had a minimum at order 4. Both measures will coincide if the estimated variance is exactly the true variance, or if all estimated parameters of the AR model are very close to their theoretical values. Similar results are found in a comparison of the spectral distortion of (12) or IRSE of (13) and their variants with $h(\omega)$ that include the estimated variance. The performance in this figure is in line with the poor results of IMSE in other examples.

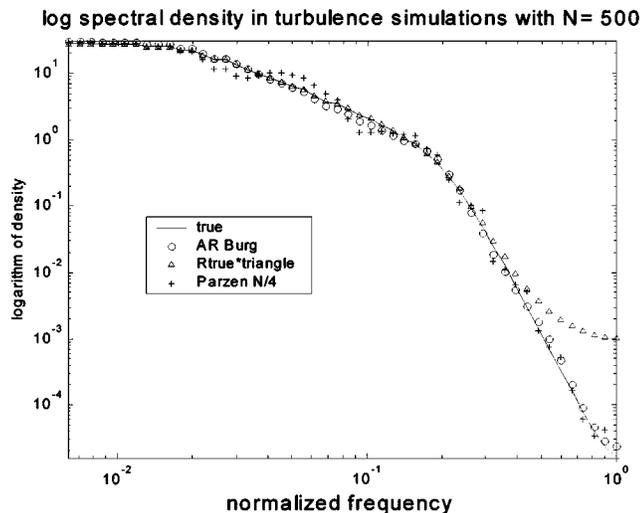


Fig. 4. True spectrum, estimated AR spectrum, true spectrum with triangular bias, and estimated tapered periodogram with Parzen window of length 125. Estimates from 500 simulated equidistant turbulence observations with spectrum (14).

IV. APPLICATION TO A TURBULENCE EXAMPLE

A second example uses a spectrum that is derived from turbulence theory [10]. It has a slope of $-5/3$ for low frequencies and a slope -7 for high frequencies

$$h(\omega) = \frac{c}{\left(1 + \frac{\omega}{\omega_1}\right)^{5/3} \left(1 + \frac{\omega}{\omega_2}\right)^{16/3}} \quad (14)$$

with $\omega_1 = 0.01\pi$ and $\omega_2 = 0.1\pi$. This spectrum has been generated with an AR(14) process that gives a close approximation to (14). Spectra can be estimated with time series methods and with modified periodograms [9], [14]. Fig. 4 shows that the spectrum of the AR model estimated with Burg's method is very close to the true spectrum in this log-log plot. The periodogram estimated with the Parzen window has a higher variance. The Fourier transform of the *true* covariance $R(i)$, multiplied with the triangle $(1 - i/N)$ has a severe bias in the higher frequencies.

Some quality measures for this turbulence example are given in Table I. The most important conclusion of this table is that IMSE is very small for the true covariance multiplied with the triangle. But the spectrum in Fig. 4 and the value for the measure ME in Table I are poor on the same row of R_{true} multiplied with the triangle $1 - i/N$. If the spectrum is plotted on a linear scale, then the appearance of the true covariance with triangle would seem reasonable because all contributions above the normalized frequency 0.2 would be nearly zero. Small values for IMSE are obtained if errors are small in the strong part of the spectrum. Large relative errors in the weak part of the spectrum have a negligible influence on the measure IMSE. The absolute measures IMSE and RED show a remarkable difference in Table I for the triangular distortion.

The three measures ME, SD, and IRSE of the relative class were nearly identical for $N = 500$, as they should be according to the asymptotic theory. Estimating from more observations gives a closer relation between the different criteria in this class. The three criteria reach a minimum at order 6 and the quality of

TABLE I
THREE QUALITY MEASURES FOR THE
TURBULENCE SPECTRUM OF FIG. 4. ESTIMATED WITH DIFFERENT TIME SERIES
MODELS, WITH A MODIFIED PERIODOGRAM AND THE INFLUENCE OF THE
TRIANGULAR BIAS ON THE THEORETICAL COVARIANCE FUNCTION

	ME	IMSE	RED
AR	9.3	0.056	.006
MA	71.8	5.502	.320
ARMA	9.6	0.250	.012
$R_{\text{true}} * \text{Triangle}$	569.3	0.005	.033
Parzen N/4	42.4	0.424	.296

Measures for AR models in turbulence simulations with N= 500

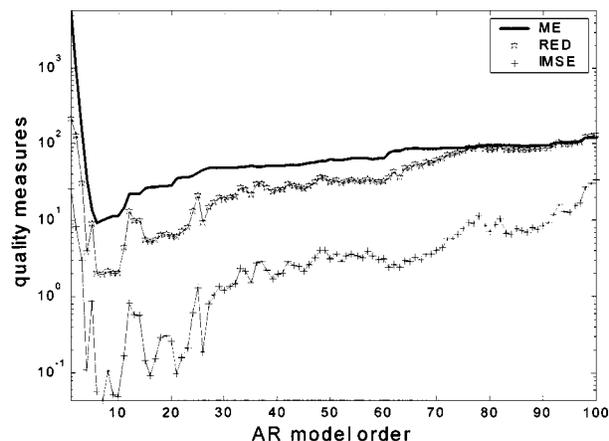


Fig. 5. Three types of quality measures as a function of the order of the AR model for 500 observations of a turbulence process.

higher order models is monotonically increasing, as might be expected. As in Fig. 1, the quality of AR models as a function of the AR order has been studied for different measures for this example in Fig. 5. The behavior of IMSE as a function of the model order is irregular, especially for AR orders between 6 and 27. No reason has been found in the AR models that justifies this behavior. Obviously, IMSE is not a good measure for the mutual quality of those AR models estimated from the same data.

The insensitivity of IMSE to frequencies with a small spectral density can be demonstrated by calculating the effect of truncating the estimated spectrum, i.e., by replacing the high-frequency part of the estimated spectrum by zero

$$\hat{f}_\alpha(\omega) = \begin{cases} \hat{f}(\omega), & |\omega| \leq \alpha\pi \\ 0, & \alpha\pi < |\omega| \leq \pi. \end{cases} \quad (15)$$

This truncation represents an important change in the character of the process [9, p. 733]. The relative measure ME has the estimated spectrum in the denominator, see (7) and (10), and the quotient equals ∞ if an estimate zero is substituted. The ME of this truncated spectrum equals ∞ for all truncation points α smaller than 1. Table II shows that IMSE with the truncated spectrum $\hat{f}_\alpha(\omega)$ does not change in the 4th decimal place for $\alpha > 0.4$ for all three estimated spectra. So for the second half of this example spectrum, taking the estimated spectral density or taking zero instead has no consequences for the numerical value of IMSE. This is the same for all three estimates in this example

TABLE II
IMSE OF TRUNCATED SPECTRA $\hat{f}_\alpha(\omega)$ AS A FUNCTION OF THE TRUNCATION POINT α FOR AN AR SPECTRUM, A WINDOWED PERIODOGRAM, AND FOR THE THEORETICAL SPECTRUM WITH TRIANGULAR BIAS

α	AR	Parzen	triangle
.004	9.7441	9.7302	9.7287
.008	7.2809	7.2580	7.2569
.012	5.3911	5.3645	5.3607
.02	3.0302	3.1019	2.9938
.04	0.9833	1.2068	0.9457
.08	0.2496	0.6008	0.2042
.12	0.1066	0.4708	0.0559
.16	0.0679	0.4345	0.0170
.2	0.0586	0.4258	0.0075
.3	0.0565	0.4239	0.0054
.4	0.0565	0.4238	0.0054
1	0.0565	0.4238	0.0054

because the relative spectral density $f(\alpha)/f(0)$ is less than 0.01 for $\alpha > 0.4$, and the numerical value of IMSE shows only variations in the fifth decimal. Hence, the character of the model can become completely different without a noticeable change in the quality criterion IMSE.

V. CONCLUDING REMARKS

The class of relative measures with the model error ME, the prediction error, the likelihood ratio or the spectral distortion has a sound mathematical background as a measure for the spectral quality in a mutual evaluation of estimation algorithms. This class has also a sensible practical meaning given by the prediction error in the time domain. These measures give reliable conclusions if the general shape of spectral models is important. They all represent a relative measure in the frequency domain. The model error ME is the preferred normalized version in this class that is independent of true and estimated variances. The spectral distortion with variance can be used if the estimated variance is also important for the quality.

The integrated squared measure IMSE has no firm theoretical background and can give undesirable results in practice. Models with a small value for IMSE can be very inaccurate at frequencies with little power. The RED is useful in speech but not in general applications.

REFERENCES

- [1] A. H. Gray and J. D. Markel, "Distance measures for speech processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, no. 5, pp. 380–391, 1976.
- [2] J. S. Erkelens and P. M. T. Broersen, "Equivalent distortion measures for quantization of LPC model," *Electron. Lett.*, vol. 31, no. 17, pp. 1410–1412, 1995.
- [3] P. M. T. Broersen, "The quality of models for ARMA processes," *IEEE Trans. Signal Processing*, vol. 46, pp. 1749–1752, June 1998.
- [4] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
- [5] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*. Uppsala, Sweden: Almqvist and Wiksells Boktryckeri AB, 1956.
- [6] E. Parzen, "Some recent advances in time series modeling," *IEEE Trans. Automat. Contr.*, vol. AC-19, no. 6, pp. 723–730, 1974.
- [7] T. Sundin and P. Stoica, "On nonparametric spectral estimation," in *Signal Processing IX, Theories and Applications, Proc. Eusipco*, vol. 1, 1998, pp. 73–76.
- [8] B. Friedlander and B. Porat, "A general lower bound for parametric spectrum estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, no. 4, pp. 728–733, 1984.
- [9] M. B. Priestley, *Spectral Analysis and Time Series*. New York: Academic, 1981.
- [10] E. Müller, H. Nobach, and C. Tropea, "Model parameter estimation from nonequidistant sampled data sets at low data rates," *Meas. Sci. Technol.*, vol. 9, pp. 435–441, 1998.
- [11] H. R. E. van Maanen and A. Oldenziel, "Estimation of turbulence power spectra from randomly sampled data by curve-fit to the autocorrelation function applied to laser-doppler anemometry," *Meas. Sci. Technol.*, vol. 9, no. 15, pp. 458–467, 1998.
- [12] J. S. Erkelens and P. M. T. Broersen, "Reconstruction error distortion measure for quantization of LPC models," *Electron. Lett.*, vol. 32, no. 15, pp. 1347–1349, 1996.
- [13] 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.
- [14] —, "Facts and fiction in spectral analysis," *IEEE Trans. Instrum. Meas.*, vol. 49, pp. 766–772, Aug. 2000.
- [15] G. A. F. Seber, *Linear Regression Analysis*. New York: Wiley, 1977.
- [16] S. de Waele and P. M. T. Broersen, "Error measures for resampled irregular data," *IEEE Trans. Instrum. Meas.*, vol. 49, pp. 216–222, Apr. 2000.
- [17] G. Wilson, "Factorization of the covariance generating function of a pure moving average process," *SIAM J. Numer. Anal.*, vol. 6, pp. 1–7, 1969.

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. His main research interest is automatic identification. 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 will be the automatic identification of input-output relations.