Spectral Analysis of EEG's by Autoregressive Decomposition of Time Series *

WILL GERSCH†
School of Aeronautics, Astronautics, and Engineering
Purdue University, Lafayette, Indiana

Communicated by K. Åström

ABSTRACT

A technique for scalar and multidimensional spectral analysis based on the autoregressive representation of the observed data records is presented and illustrated. In an autoregressive representation the observed data set is regressed on its own past history. This results in a formula that expresses the observed data as the output of a linear filter excited by an uncorrelated sequence ("discrete white noise"). Energy spectral densities, transfer functions, and coherences are computed from the autoregressive formula. The results of this technique are compared with the older windowed periodogram methods of spectral analysis. Two potential advantages over the latter methods are observed. For spectral estimates of comparable statistical performance, the autoregressive method analysis appear smoother and easier to interpret than the older windowed periodogram analysis. Also, in contrast with the expertise required to apply the windowed periodogram analysis, use of the autoregressive representation spectral analysis method appears to require little or no subjective judgment.

INTRODUCTION

In this paper an autoregressive representation method for computing the energy spectral density of single and multiple EEG test series and transfer function and coherence between pairs of EEG time series is communicated and illustrated. Results are compared with the older

* The work reported here was supported by NIH Special Training and Research Fellowship 1F 10 NB 1686-C1A, NSRB at Stanford University, School of Medicine, Division of Neurology. The computations were performed at, and supported by, the ACME facility of the Stanford University Medical Center.

† Present address: Stanford University School of Medicine, Stanford, California.

Mathematical Biosciences 7 (1970), 205–222

"windowed" periodogram spectral analysis procedure. In the autoregressive-representation procedure the observed time series (record or set of records) is expressed in a formula (best in a least squares sense) in terms of its own past history plus an additional unpredictable, uncorrelated random component. In this sense, the observed record data are regressed (projected) on themselves. The autoregressive data series formula is an explicit parametric model of the observed data set from which spectral densities, coherences, and transfer functions between time series may be computed. These formulas and illustrations of their meaning and use are presented here. The technical question of how long an autoregressive model to fit to the data is determined by a decision-theoretic procedure that is equivalent to determining the number of regressors in a conventional regression analysis.

A windowed periodogram spectral analysis technique was communicated in the neurophysiological literature in 1963 [10]. Dating from approximately that time, that technique and slight variations of it have been systematically employed by researchers in neurophysiology (see literature review by Matousek [8]). At approximately the same time, Whittle [11], demonstrated that all lengths of fit of autoregressive models to arbitrary (stationary time series) observed data were stable. Jones in 1966 [6] used autoregressive representations for time series predictions. Since then, important theoretical contributions on autoregressive representation spectral analysis by Parzen [9], Akaike [1], and Kromer [7] have appeared.

This paper contains formulas, computational details, and critical comments on the application of the autoregressive representation of multiple times series for the computation of energy spectral densities, coherences, and transfer functions. Illustrative examples compare the statistical performance of this technique with the windowed periodogram analysis of electroencephalographic (EEG) data and on a known model.

METHODS

The observed data set, a finite duration single or multichannel continuous record, is sampled at regular intervals and the data are stored in the memory of a digital computer. Computations are performed to determine a formula that explains the observed time series in terms of its own past history plus an additional unpredictable component. The energy spectral densities of the individual data channels, transfer function, and coherence between channels are computed from that formula.

The mathematical details of the computations are described in the Appendix.

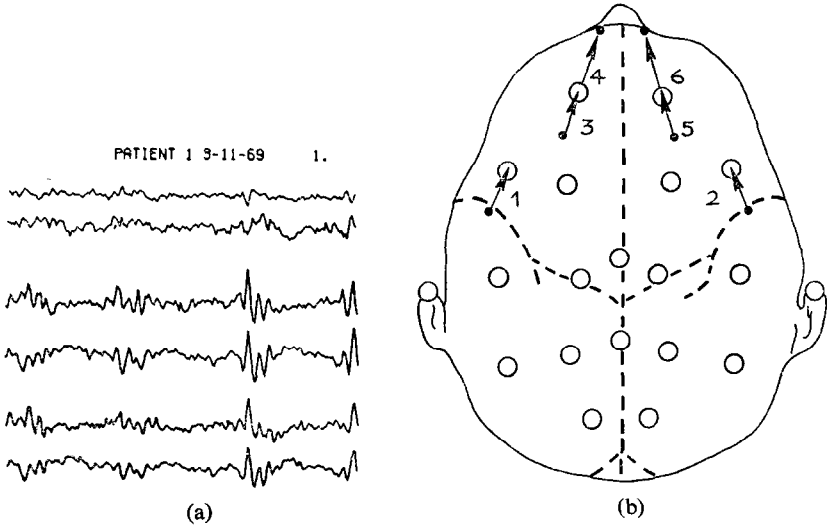


FIG. 1. Six Channels of human scalp EEG data. (a) 8-sec intervals of 6 channels of continuous EEG records, channels numbered sequentially 1 to 6 from top to bottom. (b) Scalp placement of bipolar electrodes.

RESULTS

Figure 1 illustrates the six channels of simultaneously recorded bipolar channels of EEG data taken from the scalp of a human epileptic patient that are used for data analysis in this paper. The 8-sec record was sampled regularly at 10 msec intervals and stored in computer memory. The

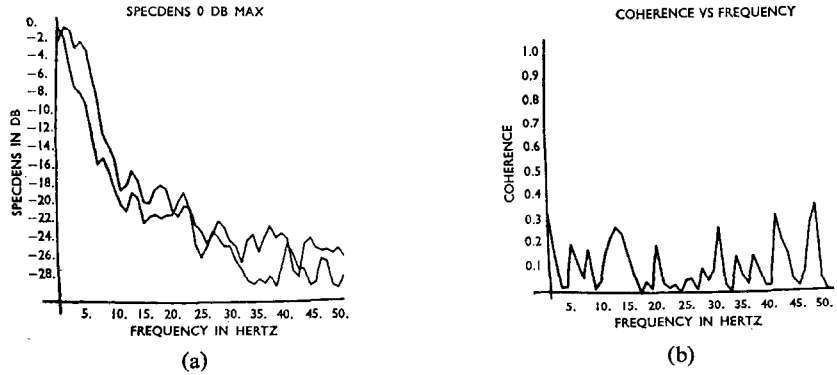
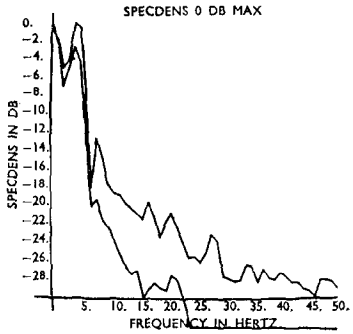
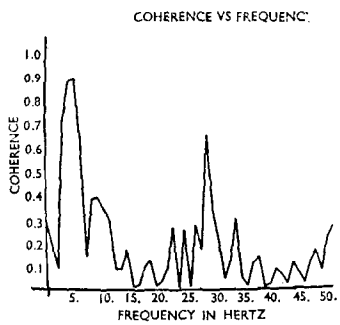


FIG. 2. Windowed periodogram spectral computations. (a, b) Data for Fig. 1a, channels 1, 2. (c, d) Data for Fig. 1a, channels 3, 4.

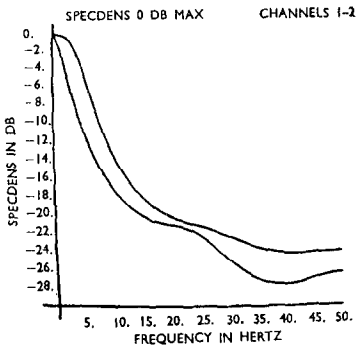


(c)

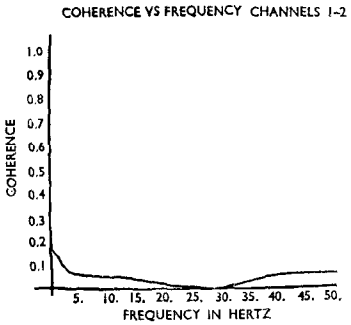


(d)

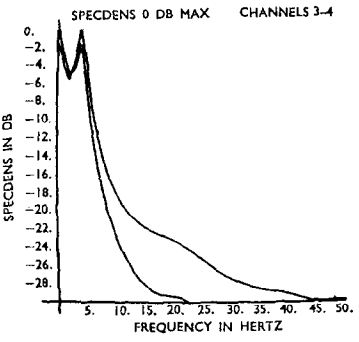
FIG. 2 (Continued).



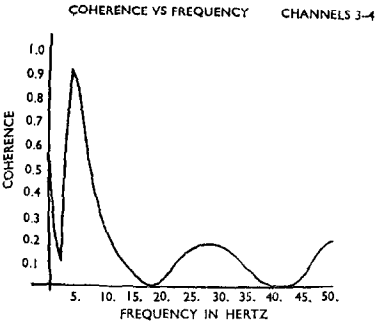
(a)



(b)



(c)



(d)

FIG. 3. Autoregressive representation spectral computations. (a, b) Data for Fig. 1a, channels 1, 2. (c, d) Data for Fig. 1a, channels 3, 4.

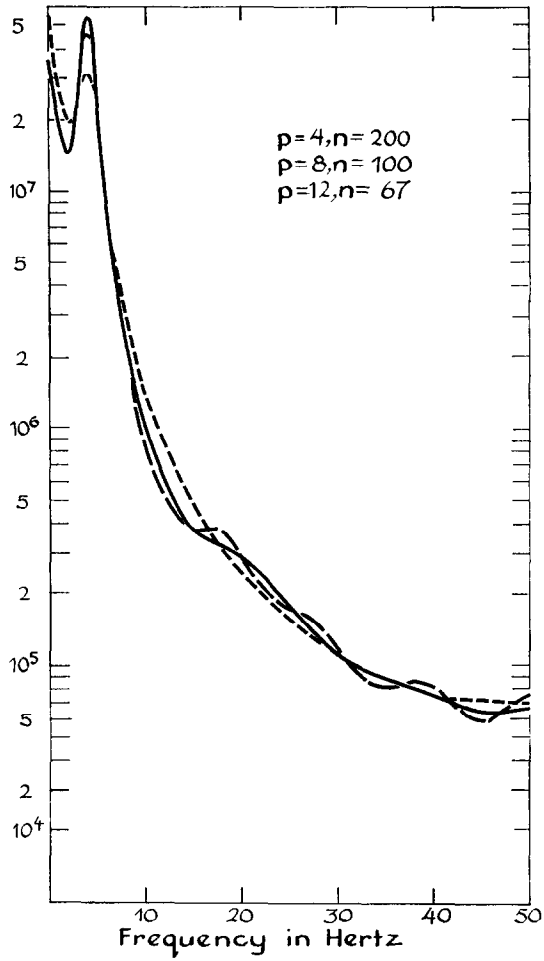


FIG. 3 (Continued). (e) Spectral density, channel 3, parametric in p.

"spikes" in the data record, (Fig. 1) are epileptic spikes (Dr. Barry Tharp, Stanford University Medical Center provided the data).

In Fig. 2, energy spectral density and coherence results computed from the data in channels 1–4 in Fig. 1a using a windowed periodogram spectral analysis technique (the Akaike trigonometric window W2, with bandwidth $b = 1$ Hz and 16 degrees of freedom [2], are shown. Figure 3 illustrates results, computed over the same data used to compute Fig. 2, determined by the autoregressive representation technique. Figures 3a, 3b and 3c, 3d were computed with 160 and 133 degrees of freedom, respectively.

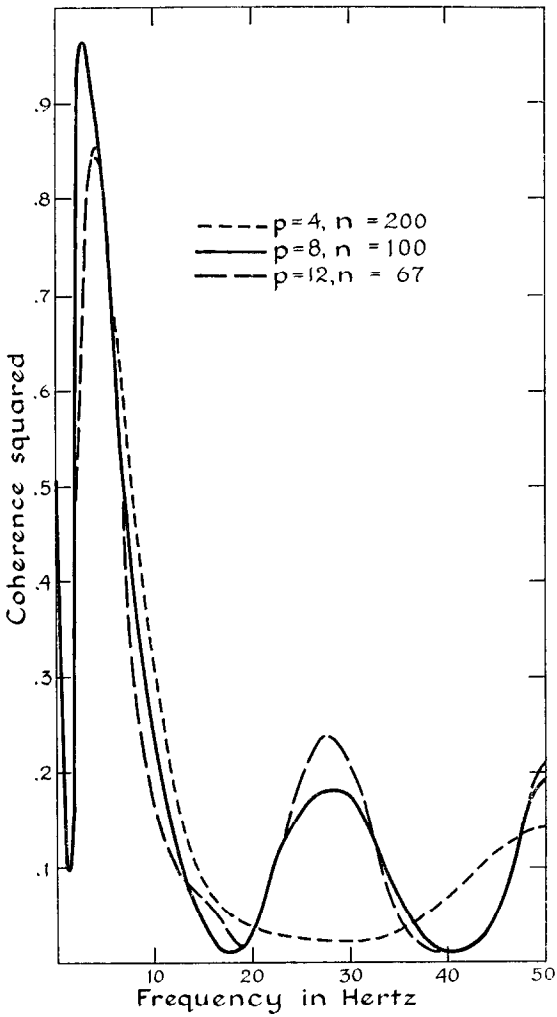


FIG. 3 (Continued). (f) Spectral coherence, channels 3, 4, parametric in p.

Figure 4 illustrates the comparative performance of the autoregressive representation method and the windowed periodogram method of spectral analysis on a known model.

DISCUSSION

An autoregressive (a.r.) representation of a multiple time series is formally equivalent to considering that time series to be generated by

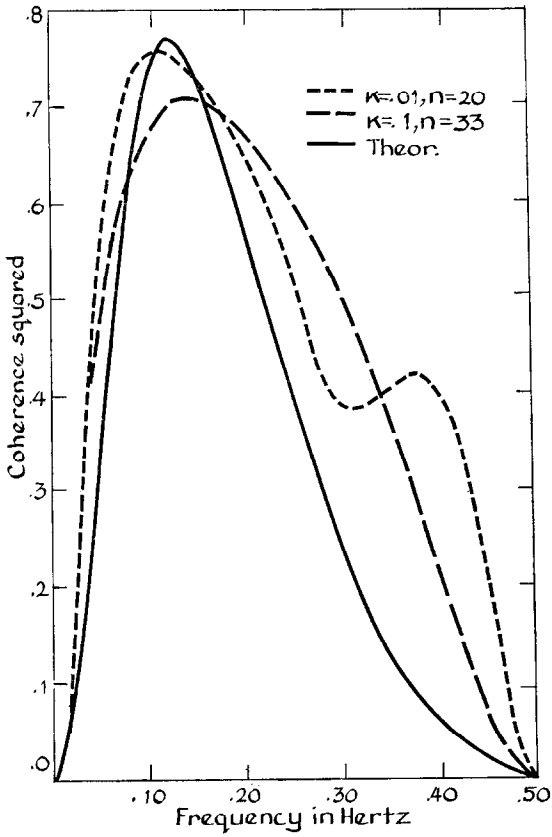


FIG. 4. Spectral coherence computations on a known model. (a) Autoregressive analysis, 100 data points for $k = 0.01$ and $k = 0.1$.

passing white noise through a linear filter (multidimensional), and determining the filter characteristics and the statistical characteristics of the white noise. The autoregressive representation is an answer to the question: What are the parameters of the system model that generated the observed data? The spectral computations, energy spectral density, transfer function, and coherence are well-defined attributes of that system model.

Windowed periodogram (w.p.) spectral analysis is the windowed or smoothed periodogram. The latter is equivalent to the fourier transform of the sample record covariance function. Self- and cross-spectral densities are computed at each frequency. Transfer function and coherence are well-defined entities computed from those estimates.

The objective of the smoothing in the w.p. method is to improve the

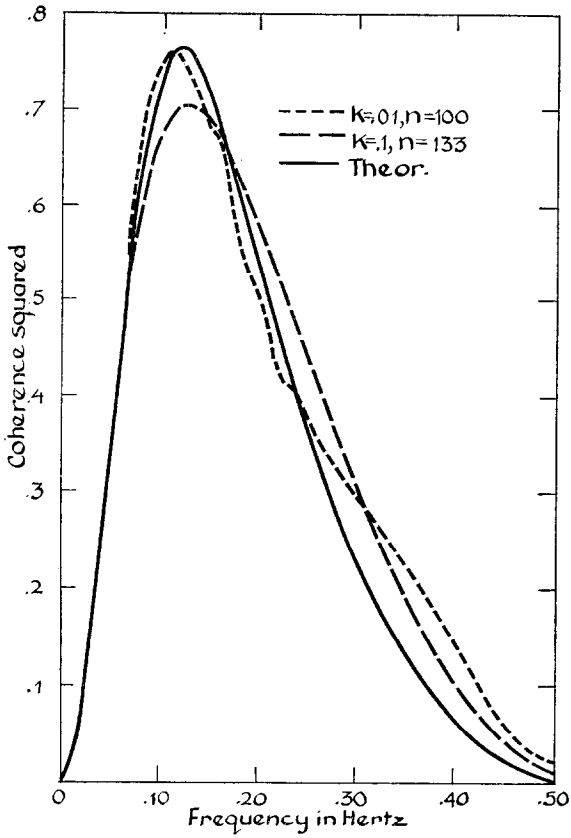


FIG. 4 (Continued). (b) Autoregressive analysis, 400 data points for $k = 0.01$ and $k = 0.1$.

statistical properties of the unsmoothed estimates. Increasing the smoothing (by widening the spectral window) decreases the variance but increases the bias or distortion of the spectral estimates. Because the spectral density of the record in question is not known *a priori*, it is not possible to design a spectral window that optimally distributes bias and variance spectral estimate errors. The spectral windows designed by Akaike, Daniell, Hamming, Parzen, and Tukey, for example, each distribute the bias and variance errors differently. The practice of spectral analysis is quite empirical. The window-closing technique described by Jenkins and Watts [5] is one acceptable approach. In this procedure, for a particular spectral window, the spectral window bandwidth is initially chosen to be wide (corresponding to a large number of degrees of freedom and a small

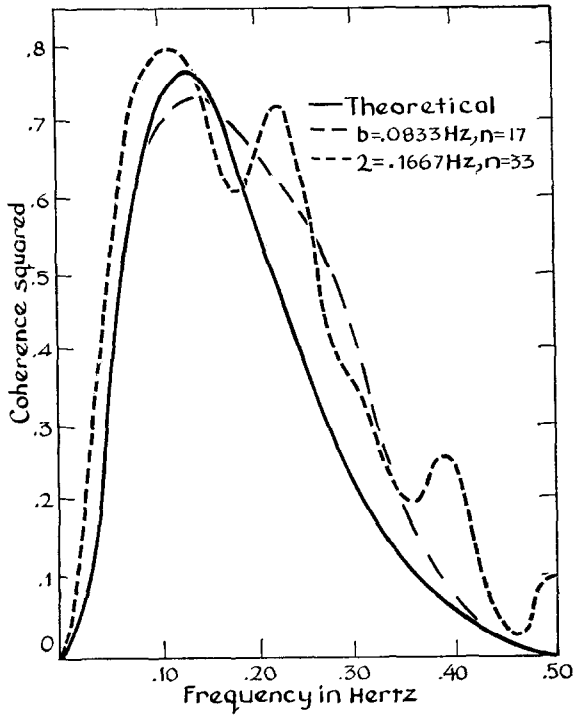


FIG. 4 (Continued). (c) Windowed periodogram analysis Tukey window, 100 data points, for bandwidths 0.0833 Hz and 0.1667 Hz.

variance), so that the general shape of the spectrum can be determined. The wide bandwidth and relatively large bias errors may mask some of the detail in the spectrum. More detail can be explored by narrowing the spectral window (or equivalently by decreasing the spectral bandwidth). As the bandwidth is narrowed, however, spurious peaks appear almost anywhere. They are a consequence of the increased statistical variability of the spectral estimates. Chapter 7 of [5] describes some informative examples of spectral computations computed from simulations of known models for different windows that illustrate the trade-off between fidelity or resolution of detail and spurious bumpiness.

The expertise required to practice the w.p. method of spectral analysis reliably is confined in the a.r. method to the determination of the length p of the autoregressive model used to represent the observed data. We employ a decision-theoretic procedure to specify p that appears to be highly successful. The a.r. method formula accounts for the energy in the

observed record. In our procedure, starting at $p = 1$, the value of p is successively increased until the relative difference between a measure of the preceding and current estimates of the observed energy that cannot be accounted for is less than a prescribed number k. Our procedure for specifying the a.r. formula length p is equivalent to Anderson's multi-alternative hypothesis test procedure for determining the number of regressors in a conventional regression analysis [4]. (See Appendix A2, Eqs. (A2.7), (A2.8).)

Figure 2 illustrates the energy spectral density and coherence computed from the data in channels 1–4 in Fig. 1a using the Akaike W2 window in the classical w.p. method of spectral analysis. The results illustrated are for a spectral bandwidth of 1 Hz with 16 degrees of freedom. The W2 window has the property that it is bias-free for local odd and quadratic variations in the spectral density. The appearance of the spectral density computed on these data using Akaike window W1 (bandwidth 1.33 Hz, 21 degrees of freedom) bias-free for odd-order variations in the local spectral density, was very similar. The computations were also performed on a 16-sec data interval (the first half of which included the 8-sec record shown in Fig. 1a) in a way that doubled the number of degrees of freedom in the computations. The results were very similar to those illustrated in Fig. 2, suggesting that the records are stationary and that the resolution and computational accuracy are reasonable. Figure 2 is therefore a studied estimate of the spectral content of the specified data records.

Comparison of Figs. 2 and 3 indicates similarity in the structural detail of the spectral estimates. These illustrations, therefore, suggest that the performance of both the a.r. and w.p. methods is comparable in this example. One advantage of the a.r. representation spectral analysis technique over the windowed periodogram technique is immediately suggested. The relative smoothness of the appearance of the a.r. representation computations (Fig. 3) allows for much more easily interpreted results than are available from the relatively "bumpy" windowed periodogram analysis (Fig. 2). For example, the "no coherence" conclusion between channels 1 and 2 compared with the distinct coherence between channels 3 and 4 is a conclusion that is more easily drawn from the a.r. representation computations than from the windowed periodogram computations.

A detailed analysis of the statistical performance of the a.r. method is beyond the scope of this paper. Parzen [9] conjectured that n times the spectral matrix is distributed with a complex Wishart distribution with $n = N/2p$ degrees of freedom and that the a.r. cross-spectral estimators have much smaller bias than the corresponding w.p. method estimators

(N is the number of samples in each component of the observed vector times series). Therefore, using a known theorem, on the marginal distribution of multidimensional Wishart distribution [4] $n = N/p$ is the number of degrees of freedom of each spectral estimate. Akaike [1] and Kromer [7] demonstrated that this result holds for a scalar time series, and in a simulation study we have obtained excellent support for the Parzen conjecture for two- and three- dimensional times series. Also, Kromer demonstrated that the a.r. spectral estimates are asymptotically normal, consistent, and efficient in decrease of bias and variance with sample size. That is, the asymptotic bias and variance performance of the a.r. method is superior to that achieved by any particular choice of window in the w.p. method.

In the w.p. method the notion of bandwidth b, number of degrees of freedom n, and the variance of the spectral estimate $\hat{S}(f)$ are related by the formulas [5]

$$b = \frac{n}{2T}, \qquad \text{Var } \hat{S}(f) = \frac{2}{n} S(f)^2,$$

$$\text{Var } \hat{S}(f) \times b = \frac{S(f)^2}{T} = \text{constant},$$

where T is the record duration in seconds and $S(f)$ is the true value of the spectral density at frequency f. The notion of bandwidth of a spectral window, in the w.p. method, is associated with the "width" of a peak in the spectrum that can be distinguished in the computations. Indeed, the width of the narrower peaks appearing in Fig. 2c are about 1 Hz wide. Notice that the a.r. method spectral estimates in Fig. 3c clearly resolve the principal peaks in the low-frequency region, but that the bandwidth specified by the formulas above is 8.3 Hz for this figure. The resolvability of the a.r. method is apparently comparable with that of the w.p. method; however, the concept of bandwidth, familiar in the w.p. method, does not translate readily to the a.r. method. We do note in passing that, for a scalar first-order autoregressive time series ($p = 1$), the spectral "width" is only a function of the single autoregressive parameter and that this width can be arbitrarily narrow (see [9, Chap. 7] for some examples). A more complete treatment of spectral resolvability is deferred for later consideration.

It is interesting to illustrate the sensitivity of the spectral estimates to variations in p, the length of the a.r. scheme representation, and the sensitivity of the value p to the decision theoretic procedure.

Figures 3e, 3f illustrate the spectral density of channel 3 and the coherence between channels 3 and 4 for $p = 4, 8, 12$ ($n = 200, 100, 67$). The results are quite compatible with the indicated values of n.

Table I lists values of V_p, the matrix of residuals ($\times 10^{-5}$) for $p = 0$,

TABLE I

V_0	V_6	V_{12}	V_{18}
$\begin{bmatrix} 3.642 & 3.248 \\ 3.248 & 2.064 \end{bmatrix}$;	$\begin{bmatrix} 0.3192 & 0.1354 \\ 0.1354 & 0.5731 \end{bmatrix}$;	$\begin{bmatrix} 0.3148 & 0.1327 \\ 0.1327 & 0.5591 \end{bmatrix}$;	$\begin{bmatrix} 0.3066 & 0.1289 \\ 0.1289 & 0.5529 \end{bmatrix}$

6, 12, and 18 for the data of channels 3 and 4. V_p is an absolute measure of the "unexplained" or residual energy in the observed time series. Table II lists k versus p, $p = 1, \ldots, 8$ (k is the relative measure of the difference in residual energy for successive values of p).

$k\,(p = 1)$ is always 1, because there is no predecessor to V_0 (see Eq. (A2.8)). The first two regressive coefficient matrices A_1 and A_2 account for a large proportion of the observed energy and, as indicated by the values of V_p and $k(p)$, continued increase in p only slightly decreases the residuals. In our decision procedure we determine the value p for which $k \leqslant 0.05$ and then choose $p = p + 1$. Correspondingly, Fig. 3c and 3d were computed for $p = 6$. Note that $k \leqslant 0.05$ for $p \geqslant 5$ and that spectral computations for $6 \leqslant p \leqslant 12$ do not vary very greatly.

It is desirable to provide additional documentation to demonstrate that the smoothing achieved by the a.r. representation is not arbitrary and that the statistical performance of a.r. spectral estimates is at least comparable to that achieved by the w.p. method. This we do pictorially by examining computational results on a known model and comparing these results with those obtained by the w.p. method.

Figures 4a and 4b illustrate a.r. spectral coherence estimates for realizations of $N = 100$ and 400 data points of the simulation of a known model [5, Eq. 8.1.20]. Results determined for $k = 0.1$ and $k = 0.01$ and the theoretical model data are shown. Figure 4c (from [5]) illustrates results achieved using the Tukey windowed w.p. method with $N = 100$ and two different values of bandwidth $b = 0.0833$ Hz and 0.1667 Hz and $n = 17$ and 33, respectively. Figure 4b, $N = 400$, indicates little difference between spectral coherency estimates achieved for $k = 0.1$ and $k = 0.01$. The respective degrees of freedom are 133 and 100, and the results are

TABLE II

	p							
	1	2	3	4	5	6	7	8
k	1.00	173	0.851	0.121	0.029	0.033	0.022	0.010

quite indicative of the theoretical results. (Spectral coherency estimates for $N = 800$ are identical for $k = 0.1$ and $k = 0.01$) The a.r. computed coherence estimates of Fig. 4a for $N = 100$, $k = 0.1$ and $k = 0.01$ ($n = 20$ and 33) are generally comparable, respectively, to the w.p. excessive variance-spurious bumps ($b = 0.0833$, $n = 17$), excessive bias ($b = 0.1667$, $n = 33$) results of Fig. 4c. Further illustrations of w.p. spectral estimates on known models using Tukey and Parzen windows, parametric in N and bandwidth, are illustrated in [5, Chap. 7]. For low bias and $N = 400$ those results are quite bumpy.

On the basis of computational experiments similar to those illustrated in Figs. 2 to 4 we have adopted the value of $k = 0.05$ for EEG data analysis. Typically our results have been that values of $p \leqslant 10$ satisfy this criterion on 5 to 8-sec duration data records sampled at 5 to 10-msec intervals. The resolution of our computations appears quite satisfactory, and the resultant large value of n, the number of degrees of freedom, provide somewhat superior statistical performance to that achieved by conventional w.p. methods for the same record durations. Also, the a.r. representation method appears to provide generally smoother and more easily interpretable results than the w.p. method. Finally, as illustrated here, in contrast with the expertise required in using the w.p. method, little or no subjective judgment is required in applying the a.r. representation method. Discussions of the extensions of a.r. representations to the analysis of partial spectral coherences, multiple coherency spectra, their statistical performance, and other detailed applications are deferred to future publications.

SUMMARY AND CONCLUSIONS

An autoregressive representation of multiple time series has been applied to compute the spectral densities and coherences on simultaneously recorded EEG data. Insight into the statistical performance of the autoregressive method of spectral analysis is provided by an empirical study and by comparison with the performance of the classical windowed periodogram method on human EEG data and on data simulated from a known model.

The length of the autoregressive model that is fit to the observed data is specified by a decision procedure that appears to require little or no subjective intervention. The procedure is equivalent to determining the number of regressors in a conventional regression analysis. The number of degrees of freedom in the computation is N/p, where N is the number of observed samples (in each record of a multidimensional time series) and p is the length or order of the autoregressive model fit to the data. On

multichannel EEG record durations of 5 to 10 sec sampled at 5 to 10-msec intervals, values of $p \leqslant 10$ are generally observed. The resolution of our computations appears quite satisfactory, and the resultant large number of degrees of freedom provides statistical performance somewhat superior to that achieved by conventional windowed periodogram methods for the same record durations. Also, the autoregressive representation method appears to provide generally smoother and more easily interpretable results than the windowed periodogram method.

APPENDIX

All the data records we consider analyzing are time-continuous records. However, all the computations we do are performed on a digital computer using regularly sampled values of these data. The following discussion restricts its attention to the analysis of time-discrete data sets obtained from those sampled records. In many places the discussion is necessarily brief. The book by Jenkins and Watts [5] is an excellent background reference.

A1. ONE-DIMENSIONAL (SCALAR) TIME SERIES

Let y_t, $t = 1, 2, \ldots, N$ designate N sample values of a scalar, (one-dimensional) time series. The observed time series is assumed to be a mean-zero stationary covariance series. The autoregressive representation of the observed time series y_t is

$$y_t + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p} = x_t, \tag{A1.1}$$

where x_t is an uncorrelated sequence (the so-called discrete-time "white noise"), and p is a parameter determined from the data set y_t.

For convenience, (A1.1) is rewritten in the form

$$\sum_{i=0}^{p} a_i y_{t-i} = x_t, \qquad a_0 = 1. \tag{A1.1'}$$

A spectral-domain input-output relationship between the series y_t and x_t, $t = 1, 2, \ldots, N$ can be realized by applying Z transform theory (see [5, Chap. 2]) to (A1.1'). The Z transform of (A1.1') is

$$\sum_{k=0}^{p} a_k z^k Y(z) = X(z), \tag{A1.2}$$

where $z = \exp(-i\omega)$. The sampled record y_t has a frequency representation that is unique up to $|\omega| \leqslant \pi$ (and periodic for $|\omega| \geqslant \pi$). With the sampling interval specified as ts, $\omega = 2\pi f \cdot ts$. From the well-known

sampling theorem, the maximum frequency of representation $f_{\max}$ is $f_{\max} = 1/2ts$. Therefore $\omega = \pi f/f_{\max}$ is the formula by which information in the z domain from regularly sampled records can be transformed to a continuous frequency interpretation.

The transfer function between the two time series y_t and x_t, $H(z)$, is given by

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1}{\sum_{k=0}^{p} a_k z^k} \tag{A1.3}$$

For the case in which x_t is an uncorrelated sequence, with second moment σ^2, a well-known result in the analysis of stationary processes is that the energy spectral density $S(z)$ of the time series y_t is given by

$$S(z) = \sigma^2 H(z) H(z^{-1}) \tag{A1.4}$$

Substitution of (A1.3) in (A1.4) yields an expression for the energy spectral density of the observed time series y_t in terms of regression parameters $a_1, \ldots, a_p$. At frequency f that spectral density is

$$S(f) = \frac{\sigma^2}{(\sum_{k=0}^{p} a_k \cos \pi k f_{\max})^2 + (\sum_{k=1}^{p} a_k \sin \pi k f_{\max})^2}. \tag{A1.5}$$

The values $a_1, \ldots, a_p$ and v_p, the residual variance remaining in the observation set y_t, $t = 1, \ldots, N$ after p terms of regression, may be computed by the following artifice: Multiply (A1.1′) by y_{t-j} and take mathematical expectations to obtain the equations

$$\sum_{i=0}^{p} a_i^{(p)} g(j - i) = 0, \qquad a_0 = 1, j = 1, \ldots, p,$$

$$\sum_{i=0}^{p} a_i^{(p)} g(i) = v_p, \tag{A1.6}$$

where the quantity $g(k)$ is the kth sample covariance defined by

$$g(k) = \frac{1}{N} \sum_{k=1}^{N-k} y_t y_{t+k}. \tag{A1.7}$$

Equation (A1.6) is known as the Yule-Walker equation. It is linear in the unknown values $a_1, \ldots, a_p$, and it may be solved by successive elimination for the $a_1, \ldots, a_p$, thus avoiding the computation of a matrix inverse.

A2. MULTIDIMENSIONAL TIME SERIES

The generalizations in notation required to accommodate simultaneous sets of (multidimensional) time series is very simple. Small letters are used

to designate vectors and capital letters to designate matrices. The time series y_t and x_t may now be interpreted to mean the n vector time series.

$$y_t = \begin{bmatrix} y_{1_t} \\ y_{2_t} \\ \cdot \\ \cdot \\ \cdot \\ y_{n_t} \end{bmatrix}, \qquad x_t = \begin{bmatrix} x_{1_t} \\ x_{2_t} \\ \cdot \\ \cdot \\ \cdot \\ x_{n_t} \end{bmatrix}, \qquad t = 1, 2, \ldots, N. \tag{A2.1}$$

The general autoregressive formula for the vector time series case corresponding to (A1.1′) is

$$\sum_{k=0}^{p} A_k^{(p)} y_{t-k} = x_t. \tag{A2.2}$$

In (A2.2), y_t and x_t are n vectors each, A_k is an $n \times n$ matrix, $k = 0, 1, \ldots, p$, and $A_0 = I$, the n-dimensional identity matrix. The energy spectral density of the vector time series y_t is computed from the $n \times n$ matrix:

$$S(z) = \left(\sum_{k=0}^{p} A_k^{(p)} z^{-k} \right)^{-1} V_p \left(\sum_{k=0}^{p} A_k'^{(p)} z^{k} \right)^{-1}. \tag{A2.3}$$

In (A2.3), A_k' is the transpose of the matrix A_k. The transfer function between the ith and jth data channels is

$$H_{ij}(z) = \frac{S_{ij}(s)}{S_{ii}(z)}, \qquad i, j = 1, \ldots, n. \tag{A2.4}$$

In (A2.4), $S_{ij}(z)$ is the ij element of the $S(z)$ matrix.

Finally, the coherence between the ith and jth time series is

$$W_{ij}^2(z) = \frac{S_{ij}(z) S_{ij}(z^{-1})}{S_{ii}(z) S_{jj}(z)}, \qquad i, j = 1, \ldots, n. \tag{A2.5}$$

Observe that, given the values p and A_k, $k = 1, \ldots, p$ in (A2.2), Eqs. (A2.4) and (A2.5) specify the transfer function and coherence between arbitrary pairs of data sets. The values $A_k^{(p)}$, $n = 1, \ldots, p$ in (A2.2) may be computed by solving a multidimensional form of the Yule-Walker equations.

The inverse of the complex $n \times n$ matrix

$$M = \sum_{k=0}^{p} A_k^{(p)} z^{-k}$$

required in (A2.3) may be computed by the well-known trick of writing

that matrix in the form of a real $2n \times 2n$ matrix, M_{2n}:

$$M_{2n} = \begin{bmatrix} R(z) & I(z) \\ -I(z) & R(z) \end{bmatrix}, \tag{A2.6}$$

where

$$R(z) = \sum_{k=0}^{p} A_k^{(p)} \cos k\omega,$$

$$I(z) = \sum_{k=1}^{p} A_k^{(p)} \sin k\omega,$$

$R(z)$ and $I(z)$ are respectively, the $n \times n$ real and imaginary matrix parts of the M matrix.

The inverse of (A2.6) has the same structural format as M_{2n}. The inverse of

$$\sum_{k=0}^{p} A_k'^{(p)} z^k,$$

also required in (A2.3), is the complex conjugate transpose of the inverse of

$$\sum_{k=0}^{p} A_k^{(p)} z^{-k}.$$

Using that fact, only a single inverse need be computed at each frequency to compute (A2.3).

The only technical issue in using the autoregressive representation spectral analysis method is to decide how large a value of p is to be used before truncating the time series (A1.1) and (A2.2). One approach to that decision is to form the left-hand side of the autoregression from the observed data and test the residual series (the right-hand side of that equation) for "whiteness." The first p for which the residual series is white is a sufficient p. We note that whiteness tests for scalar time series are not terribly critical ([Ref. 9, Chap. 6]), and that the situation of testing for jointly white (multiple) time series is decidedly poorer. Alternatively, one approach for deciding the regression length of a scalar time series is to examine the successive residual energies $v_1, v_2, \ldots$ and terminate the procedure when a sufficient proportion of the observed energy has been explained. The formula employed for scalar time series is: Terminate the recursion at p when

$$\frac{v_{p-1} - v_p}{v_p} \leqslant k, \tag{A2.7}$$

with $k = 0.05$. With N large, the left-hand side of (A2.7) multiplied by

$(N - p)$ is approximately χ^2 with one degree of freedom [6]. The multidimensional equivalent of (A2.7) is: Terminate the recursion at p when

$$\frac{|V_{p-1}| - |V_p|}{|V_p|} \leqslant k, \tag{A2.8}$$

with $k = 0.05$. The vertical lines in (A2.8) denote the determinant of the enclosed matrix. The left-hand side of (A2.8) multiplied by $(N - pd)/d$ is approximately χ^2 with d degrees of freedom [6]. Formulas (A2.7) and (A2.8) conform to the statistical tests suggested by Anderson [4] for determining the number of regressors in single and multidimensional regression analysis.

REFERENCES

1 H. Akaike, Power spectrum estimation through autoregressive model fittings, *Inst. Stat. Math., Tokyo*, Res. Memo. No. 27, 1969 (to appear in *Ann. Inst. Stat. Math.*).

2 H. Akaike and Y. Yamanouchi, On the statistical estimation of frequency response function, *Ann. Inst. Math.*, **14**(1962), 23–56.

3 T. W. Anderson, *Introduction to Multivariate Statistical Analysis*, Wiley, New York, 1958.

4 T. W. Anderson, The choice of the degree of a polynomial regression as a multiple decision problem, *Ann. Math. Stat.*, (1) **33**(1962), 255–265.

5 G. Jenkins and D. Watts, *Spectral Analysis and Its Applications*, Holden Day, San Francisco, 1968.

6 R. H. Jones, Prediction of multivariate time series, *J. Appl. Meteorol.*, 3(1964), 285–289.

7 R. E. Kromer, doctoral dissertation, Asymptotic properties of the autoregressive spectral estimator, Dept. of Statistics, Stanford Univ., *Tech. Rept.*, No. 13, 1969.

8 M. Matousek, Automatic analysis in clinical electroencephalography, *Psychiat. Res. Inst. Prague, Res. Rep.*, No. 9, p. 240.

9 E. Parzen, Multiple time series modelling, *Multivariate Anal.*, **2**(1969), 389–409.

10 D. O. Walter, Spectral analysis for electroencephalograms, mathematical determination of neurophysiological relationships from records of limited duration, *Exptl. Neurol.*, **8**(1963), 155–181.

11 P. Whittle, On the fitting of multivariate autoregressions and the approximate canonical factorization of a spectral density matrix, *Biometrika*, **50**(1963), 129–134.