NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Energy Research and Development Administration, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

EVALUATION OF THE AUTOREGRESSION TIME-SERIES MODEL FOR ANALYSIS OF A NOISY SIGNAL^a

J. W. Allen

Oak Ridge National Laboratory, Oak Ridge Tennessee 37830

INTRODUCTION

The autoregression (AR) time-series model of a continuous noisy signal was statistically evaluated to determine quantitatively the uncertainties of the model order, the model parameters, and the model's power spectral density (PSD). The result of such a statistical evaluation enables an experimenter to decide whether an AR model can adequately represent a continuous noisy signal and be consistent with the signal's frequency spectrum, and whether it can be used for on-line monitoring.

Although evaluations of other types of signals have been reported in the literature, this author has found no direct reference to AR model's uncertainties for continuous noisy signals; yet the evaluation is necessary to decide the usefulness of AR models of typical reactor signals (e.g., neutron detector output or thermocouple output) and the potential of AR models for on-line monitoring applications. AR and other time-series models for noisy data representation are being investigated by others since such models require fewer parameters than the traditional PSD model (1-4). For this study, the AR model was selected for its simplicity and conduciveness to uncertainty analysis, and controlled laboratory bench signals were used for continuous noisy data.

MATHEMATICAL CONSIDERATIONS

The AR model is given by

$$y_t - \sum_{i=1}^p \phi_i y_{t-i} = a_t, \quad t = 1, 2, \dots, N, \quad (1)$$

where y_t is a series of data values sampled at intervals of ΔT seconds with the mean value subtracted, ϕ_i 's are the parameters of the model to be determined, a_t is a normally distributed random variable (called the residual) with a zero mean and variance of σ_a^2 , and N is the total number of data points being modeled.

The ϕ_i 's can be calculated for a p th order model by an autocorrelation function representation, called the Yule-Walker equations (5), that is,

$$\bar{\phi} = R^{-1} \bar{r},$$

where

$$\bar{\phi}^T = (\phi_1 \ \phi_2 \ \dots \ \phi_p),$$

$$\bar{r}^T = (r_1 \ r_2 \ \dots \ r_p),$$

MASTER

(2)

^aResearch sponsored by the Energy Research and Development Administration under contract with the Union Carbide Corporation.

$$R = \begin{bmatrix} r_0 & r_1 & \dots & r_{p-1} \\ r_1 & r_0 & \dots & r_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & \dots & r_0 \end{bmatrix},$$

and

$$r_i = \frac{1}{N-1} \sum_{t=1}^{N-1} y_t y_{t+i}, \quad i = 0, 1, 2, \dots, p.$$

Here, r_i is the estimate of the autocorrelation function for lag i . The autocorrelation function formulation is used often because it lends itself to a recursive solution (6) for the ϕ 's without involving a matrix inversion. The variance of the residuals for a p th order model is related to the ϕ_i and autocorrelation functions (6) as

$$\sigma_{a,p}^2 = r_0 - \sum_{i=1}^p \phi_i r_i. \quad (3)$$

The model order, p in Eq. (1), is commonly determined by Akaike's information criterion (6) (AIC) as $p = k$, such that $AIC(k)$ is a minimum for $k = 1, 2, \dots, N$. For residuals characterized by a gaussian distribution, as in this paper, this criterion is

$$AIC(k) = N \ln \sigma_{a,k}^2 + 2k. \quad (4)$$

Another way to determine the order based on frequency domain consistency is to compare the AR power spectral density (ARPSD) with a PSD computed by a fast Fourier transform using statistically equivalent data (FFTPSD). The p th order ARPSD is given by

$$|Y(j\omega)|_p^2 = \frac{\Delta T \sigma_{a,p}^2}{\left| 1 - \sum_{i=1}^p \phi_i \exp(-j\omega \Delta T) \right|^2}, \quad (5)$$

where ω is the angular frequency, and ΔT is the sampling interval. A quantitative measure of the comparison is the total normalized squared difference (NSD) defined as

$$NSD(k) = \int_0^{F_N} \left[\frac{|y(j\omega)|_k^2 - |P(j\omega)|^2}{|P(j\omega)|^2} \right]^2 d\omega, \quad (6)$$

where k is the order of the ARPSD, F_N is the Nyquist frequency, and $|P(j\omega)|^2$ is the FFTPSD. Since independent data are used for the different PSD's, the NSD will pass through a minimum as a function of k ; the k at which the minimum is reached is taken as the best order.

The uncertainty analyses employ a Taylor series truncated at the first-order terms to propagate the autocorrelation function uncertainties to the parameters. The uncertainty of the autocorrelation functions is determined by a discrete version of the continuous uncertainty, given by Bendat and Piersol (7), in Eq. (7),

$$\sigma[R(\tau)] \cong \left[\frac{1}{T} \int_{-\infty}^{\infty} [R^2(\zeta) + R(\zeta + \tau) R(\zeta - \tau)] d\zeta \right]^{1/2} \quad (7)$$

$R(\tau)$ is the continuous autocorrelation function, and T is the length of the data record. When long data records are used (i.e., T becomes large), the uncertainty can be made arbitrarily small. Thus, the approximate uncertainty of ϕ_k is given by

$$\sigma(\phi_k) \cong \left(\sum_{i=1}^P \sum_{j=1}^P \frac{\partial \phi_k}{\partial r_i} \frac{\partial \phi_k}{\partial r_j} \delta r_i \delta r_j \rho_{ij} \right)^{1/2} \quad (8)$$

where δr_i is the uncertainty of r_i .

The approximate uncertainty of the AIC criteria is

$$\sigma[AIC(k)] \cong \left\{ \sum_{i=1}^P \sum_{j=1}^P \frac{\partial}{\partial r_i} [AIC(k)] \frac{\partial}{\partial r_j} [AIC(k)] \delta r_i \delta r_j \rho_{ij} \right\}^{1/2} \quad (9)$$

and the approximate uncertainty of the ARPSD is

$$\sigma[|Y(j\omega)|^2] \cong \left\{ \sum_{i=1}^P \left[\frac{\partial}{\partial \phi_i} (|Y(j\omega)|^2) \sigma(\phi_i) \right]^2 \right\}^{1/2} \quad (10)$$

In Eqs. (8) and (9), ρ_{ij} is the Pearson correlation coefficient (8), defined as

$$\rho_{ij} = \frac{E[(r_i - \bar{r}_i)(r_j - \bar{r}_j)]}{\left\{ E[(r_i - \bar{r}_i)^2] E[(r_j - \bar{r}_j)^2] \right\}^{1/2}} \quad (11)$$

where E is the expectation operator, and $\bar{r}_i$ is the average of the autocorrelation function at lag i for multiple runs.

DATA ANALYSIS

The analog signals for analysis were created using the circuits shown in Fig. 1. With the switch in position "a," the input to the low-pass anti-alias filter is second-order, RC low-passed noise created by two RC filters in series with $R_1 C_1 = 0.01$ sec and $R_2 C_2 = 0.02$ sec. In position "b," independent, band-passed white noise is added.

During an analysis, the data were sampled at a constant rate by an analog-to-digital converter, such that the cutoff frequency of the anti-alias filter was about 0.9 the value of the Nyquist frequency. Digitized data were analyzed using Fortran programs and a minicomputer. Pearson correlation coefficients were estimated by multiple runs of 512 points each.

Fig. 1. Two analog signals analyzed.

Uncertainty Analysis

An AR model was fitted for the signal $y(\tau)$ in Fig. 1a, using autocorrelation functions calculated from 100 blocks of 512 points each sampled at 0.0091-sec intervals for a total time of 466 sec. The minimum AIC value was $-1.72337 \times 10^5 \pm 7.33 \times 10^2$, and it occurred at order 16. However, the AIC 67% (σ) uncertainty limits covered the AIC values from order 8 to at least 20 (Fig. 2). The AIC changed very little for orders of 11 or higher, and this change is much smaller than the uncertainty. The AIC order uncertainty was checked empirically for ten multiple runs of statistically independent data; the result was an average order of 22.1 ± 6.3 .

The preceding results prompted the author to test the NSD criterion [Eq. (6)]. For ten different FFTPSD's, the average order was 14.3 ± 1.1 . A plot of the NSD vs order for a typical run (Fig. 3) shows a behavior similar to that of the AIC, that is, there is little change for orders greater than 11. The main feature of the NSD criterion is that it uses the total integrated error of a signal's frequency spectrum, giving a best-order model for the ARPSD instead of for the AR time-series. Furthermore, the empirical study of the NSD indicates that it gives a lower and more consistent order than given by the AIC. However, since a theoretical study of the NSD uncertainty was not done, we cannot conclusively state that the NSD is a better criterion.

To decide on a final order, we used both the AIC and NSD criteria as guidelines (giving the NSD more weight), but on the final determination we relied on a visual comparison of the ARPSD and the FFTPSD. The orders determined by visual inspection were often lower than those given by the AIC or NSD criteria. For the case presented, an AR(12) model was chosen.

The uncertainties of the AR(12) ϕ parameters determined without the Pearson correlation coefficients (i.e., $\rho_{ij} = 0$, $i \neq j$; and $\rho_{ii} = 1$) are extremely large compared with those calculated with the coefficients (Table 1). Thus, the coefficients should be included in an error analysis. The 95% (2σ) uncertainty limits of the ARPSD of the AR(12) model were compared with an independently calculated FFTPSD; the result was that the ϕ 's and their uncertainties were reasonably verified (Fig. 4). The ϕ 's and their uncertainties were further verified by forcing 9456 independent runs of 512 points to an AR(12) model. The calculated parameters were checked to determine if they were within the 95% uncertainty limits for 512 points (column 5, Table 1). The actual percentage of each parameter that did fall within these limits (called the "parameter tracking record") varied from 94.2% for ϕ_1 to 98.5% for ϕ_6 and ϕ_9 , indicating reliable parameters and uncertainties (column 6, Table 1).

Adequacy of the AR Model

The adequacy of an AR model and its uncertainties was judged by (1) how well the ARPSD matched an accurate FFTPSD, (2) how large the ϕ uncertainties were, and (3) how well a first-order Taylor Series approximated the ϕ uncertainties. Though adequacy judgment will vary among different people, this author arbitrarily adopted the following guidelines: (1) if an ARPSD was consistently different from an FFTPSD at any section of the spectrum for a given order, that order model was inadequate; (2) if the ϕ relative uncertainties were greater than 50%, the model was inadequate; and (3) if the tracking record could not verify the first-order Taylor series uncertainty approximation, the model's uncertainties

Fig. 2. The AIC criterion and its uncertainty vs order for second-order low-passed noise ($N = 51200$).

Fig. 3. The NSD criterion vs order for second-order low-passed noise ($N = 51200$).

TABLE 1 Uncertainties for the AR(12) Model of Second-Order Low-Pass Noise (Fig. 1a) and Its Tracking Record

1	ϕ_1	$\delta\phi_1$ without Pearson coefficients	$\delta\phi_1$ with Pearson coefficients	$\delta\phi_1$ used for tracking record	tracking record (95% expected value)
1	1.497	1.025	0.006	0.06	94.2
2	-1.028	2.476	0.013	0.13	95.8
3	0.649	3.296	0.016	0.16	97.2
4	-0.497	3.591	0.017	0.17	98.1
5	0.394	3.633	0.017	0.17	98.4
6	-0.313	3.514	0.017	0.17	98.5
7	0.248	3.296	0.015	0.15	98.4
8	-0.185	3.015	0.015	0.15	98.5
9	0.126	2.674	0.014	0.14	98.3
10	-0.080	2.245	0.012	0.12	97.6
11	0.044	1.587	0.010	0.10	96.8
12	-0.018	0.636	0.005	0.05	95.4

Fig. 4. The 95% uncertainty range of the FFTPSD and twelfth-order ARPSD for second-order low-passed noise ($\Delta T = 0.0091$ sec; $NAT \approx 466$ sec).

were inadequate. Tested against these guidelines, the AR(12) model computed in the preceding section is adequate.

Further investigation of adequacy was done by increasing the spectral magnitude dynamic range by decreasing the sampling interval. The signal in Fig. 1, with the switch in position *a*, was sampled at 0.006-, 0.004-, and 0.002-sec intervals and anti-aliased at 80, 115, and 225 Hz, respectively. The spectral magnitude dynamic ranges for these sampling intervals were about $10^{3.5}$, 10^4 , and $10^{4.5}$, respectively. Reasonable models of order 12, 12, and 7, respectively, were found using the AIC and NSD criteria and visual spectral matching.

The uncertainties in the ϕ parameters generally increased as the sampling interval decreased, and for sampling intervals of 0.004 and 0.002 sec, many of the ϕ relative uncertainties became greater than 50% (Table 2). Since the autocorrelation function uncertainties did not change significantly for the different sampling intervals, the larger uncertainties are attributed to increased parameter sensitivity to the autocorrelation functions [i.e., $\partial\phi_1/\partial r_j$ in Eq. (8)].

The tracking records for the models of the three sampling intervals were reasonable (Table 3), although the tracking record of the AR(12) model for the 0.004-sec sampling interval indicated slightly higher uncertainties than those determined, and the tracking record of the AR(7) model for the 0.002-sec sampling interval indicated slightly lower uncertainties than those determined. These discrepancies can be attributed to significant second derivatives (i.e., $\partial^2\phi_1/\partial r_j^2$) that make the first-order Taylor series inaccurate, or to uncertainties in the uncertainty.

From the preceding observations, we conclude that the AR model is adequate for anti-aliased second-order low-pass noise for sampling intervals of 0.006 sec or higher (corresponding to a spectral magnitude dynamic range less than $\sim 10^{3.5}$). For smaller sampling intervals, the AR model may be inadequate because several ϕ relative uncertainties become larger than 50%, and the first-order Taylor series may be inaccurate. The effect of these uncertainties on the ARPSD is shown in Fig. 5 for the three sampling intervals. Model inadequacy for even smaller sampling intervals than 0.002 sec was evident because the autocorrelation matrix in Eq. (2) became nonpositive definite or singular.

An On-Line Application

Given an adequate AR model based on the discussions in the previous section, it became possible to determine whether the model could be used to detect changes in a given signal. This was done by calculating an adequate model and its 95% (2σ) ϕ uncertainties for a base case signal. With these ϕ 's and their uncertainties, the parameter tracking records were determined for changes in the base case signal. The signal used for this on-line test is shown in Fig. 1, with the switch in position *b*. An adequate ninth-order AR model was computed using 100 blocks of 512 points, each sampled at 0.0091-sec intervals (a total of 466 sec). The ϕ parameters and their uncertainties are listed in Table 4. The signal shown in Fig. 1b was changed by (1) increasing the magnitude of the band-passed noise (center frequency was 20 Hz), (2) shifting the band-passed noise to a center frequency of 30 Hz, (3) widening the band-pass filter, and (4) narrowing the band-pass filter. The FFTPSD's

TABLE 2 AR Model Parameters and Their Uncertainties for Three Sampling Intervals

	$\Delta T = 0.006 \text{ sec}$		$\Delta T = 0.004 \text{ sec}$		$\Delta T = 0.002 \text{ sec}$	
	ϕ_1	$\delta\phi_1/\phi_1$	ϕ_1	$\delta\phi_1/\phi_1$	ϕ_1	$\delta\phi_1/\phi_1$
1	1.718	0.005	2.020	0.024	2.321	0.196
2	-1.204	0.014	-1.708	0.080	-2.080	0.619
3	0.666	0.033	1.118	0.172	1.100	1.466
4	-0.444	0.054	-0.830	0.256	-0.502	2.618
5	0.311	0.080	0.641	0.333	0.222	3.453
6	-0.232	0.103	-0.507	0.405	-0.091	3.682
7	0.168	0.143	0.404	0.473	0.021	4.202
8	-0.112	0.196	-0.321	0.534	--	--
9	0.080	0.250	0.251	0.597	--	--
10	-0.070	0.257	-0.192	0.608	--	--
11	0.048	0.292	0.119	0.604	--	--
12	-0.016	0.375	-0.038	0.633	--	--

TABLE 3 Parameter Tracking Record for AR Models of the Low-Passed Noise for Three Sampling Intervals (~200 Cases Each)

Parameter Number	$\Delta T = 0.006 \text{ Sec}$ Parameter Tracking record (95% expected) AR(12)	$\Delta T = 0.004 \text{ Sec}$ Parameter Tracking record (95% expected) AR(12)	$\Delta T = 0.002 \text{ Sec}$ Parameter Tracking record (95% expected) AR(7)
1	88.2	93.2	95.8
2	88.6	91.7	97.2
3	92.6	89.7	99.3
4	96.2	88.8	97.9
5	97.4	88.2	98.6
6	97.8	88.3	93.7
7	97.8	85.7	83.9
8	98.0	81.8	--
9	98.0	78.0	--
10	98.2	74.7	--
11	99.0	74.5	--
12	98.4	76.0	--
Total	95.9	84.2	95.2

(a) DT=0.006-sec AR(12) MODEL

(b) DT=0.004 sec AR(12) MODEL

(c) DT=0.002-sec AR(7) MODEL

Fig. 5. The 95% uncertainty range of the FFTPSD and ARPSD for second-order low-passed noise for three sampling intervals ($N = 51200$).

TABLE 4 Parameter Tracking Record for Variations of the Signal in Fig. 1b (100 Cases Each)

i	ϕ_i	$\delta\phi_i$ for 512 pts.	Tracking Record (95%)				
			No Change	Resonance increases (Fig. 6a)	Resonance shifts (Fig. 6b)	Resonance broadens (Fig. 6c)	Resonance narrows (Fig. 6d)
1	1.56	0.05	92.0	94.0	73.0	90.0	79.0
2	-1.20	0.11	95.0	56.0	93.0	78.0	73.0
3	0.66	0.13	97.0	74.0	58.0	48.0	94.0
4	-0.23	0.14	98.0	94.0	83.0	70.0	100.0
5	0.12	0.12	97.0	95.0	98.0	98.0	99.0
6	-0.10	0.12	96.0	95.0	93.0	91.0	100.0
7	0.04	0.12	97.0	96.0	99.0	100.0	98.0
8	-0.01	0.10	98.0	98.0	100.0	99.0	100.0
9	0.01	0.05	98.0	93.0	100.0	98.0	100.0
Total	--	--	96.4	88.3	88.6	85.8	93.7

for these four signal changes are shown in Fig. 6. It was seen for all of the changed signals that the tracking record of at least one of the parameters was well below its expected 95% value (Table 4). For example, ϕ_2 had a tracking record of 56% when the magnitude of the band-passed noise was increased, and ϕ_1 had a tracking record of 79% when the band-passed noise was narrowed. Thus, the adequate ninth-order model did detect changes in the signal.

CONCLUSIONS

Based on the analysis of second-order low-passed noise, the author makes the following conclusions:

1. The uncertainties in the AIC can make it difficult to find an optimum AR model order using this criterion.
2. Limited empirical studies show that the NSD criterion gave lower and more consistent orders than did the AIC criterion, and visual inspection allowed even lower orders for adequate ARPSD models.
3. A first-order Taylor series can be a good approximation for determining AR parameter uncertainties if the Pearson correlation coefficients of the autocorrelation functions are known.
4. The AR parameter uncertainties increase with the spectral magnitude dynamic range of a signal, eventually becoming too large to adequately portray it.
5. When an adequate AR model is determined for a signal, its parameter uncertainties can be used to detect signal changes.

REFERENCES

- (1) Fukunishi, Kohyu, Diagnostic analyses of a nuclear power plant using multivariate autoregressive processes, *Nucl. Sci. Eng.* 62, 215 (February 1977).
- (2) Wu, S. M., et al., *Feasibility Study of On-line Safety Evaluation and Monitoring of Malfunctions of the Nuclear Reactor System*, EPRI NP-189 Final Report (August 1976).
- (3) Kerlin, T. W., et al., *In Situ Response Time Testing of Platinum Resistance Thermocouples*, EPRI NP-459 (January 1977).
- (4) Upadhyaya, B. R., and T. W. Kerlin, Response time testing of temperature sensors using a noise analysis method, *Trans. Amer. Nucl. Soc.* 26, 445 (1977).
- (5) Makhoul, John, Linear prediction: a tutorial review, *Proc. IEEE* 63(4) (April 1975).
- (6) Akaike, H., A new look at the statistical model identification, *IEEE Trans. Auto. Contr.* AC-19 (December 1974).
- (7) Bendat, J. S., and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, Wiley-Interscience, New York, 1970, p. 183.
- (8) Ibid., p. 62.

Fig. 6. FFTPSD's for the signal of Fig. 1b. (a) base case; (b) band-passed noise magnitude increased; (c) band-passed noise moved to a higher frequency; (d) band-passed noise width increased; and (e) band-passed noise width decreased ($\Delta T = 0.0091$; $NAT \approx 466$ sec).