Spectrum Analysis by Autoregressive Methods*
(Performance on Application to Stationary Signals)

Minoru KAMATA** and Panyarak NGAMSRITRAGUL***

In order to develop a method capable of determining the time variant spectrum of time series, various existing approaches have been investigated. Although the Fourier-based methods are superior in their computational efficiency, their inherent characteristics may sometimes limit applications. The AR method gives the best results even for small data sets. However, insufficient information is available for determining its applicability. In this report, a brief review, as well as the performance, of various AR methods applied to a certain class of stationary time series is systematically documented. The covariance method is found to be the best solution for the determination of AR coefficients, and many trials using sinusoidal data sets indicate the usefullness and applicability of AR-based spectrum analysis.

Key Words: Signal Analysis, Spectrum Analysis, Time Series Analysis, Signal Processing, AR Method

1. Introduction

It is very common for a mechanical engineer, especially one working in the field of vibration analysis, to transform time series data into the frequency domain and make use of the resultant spectrum for his analysis. Among the transformation methods, FFT is frequently adopted because of its speed and the advent of the high-performance FFT analyzer. The FFT method which is based on the DFT, however, posseses inherent restrictions for certain applications. The frequency resolution is dependent on the time width of the observed time series. These restrictions are found to be impractical for many applications. Improvement by methods such as windowing and zero padding can be achieved to a certain level. Unfortunately, not only do the inherent characteristics of DFT remain unchanged, but these tools also introduce bias.

In contrast to the FFT method, which is classified as nonparametric, many parametric methods have been proposed. The autoregressive (or AR) methods which estimate the spectrum from the autocorrelation estimates are the most popular ones. Others such as moving-average (MA), and autoregressive moving-average (ARMA) methods can be classified as their alternatives. All of these methods are very effective in extracting frequency components from relatively short time series data. They were mainly developed for use in the field of geophysics and electrical signal processing[1]. Their applications in the field of mechanical engineering are very limited. Suzuki[2] (Tokyo Metropolitan University), among a few institutions[3],[4], has been extensively conducting research on applications of these methods to structural dynamics. The main reasons for its limited applications to such field may be the lack of clear guidelines in specifying the necessary parameters (e.g. sampling rate, number of data points, and model order).

In addition to these, many methods such as short-time Fourier transform (STFT), Wigner distribution, and wavelet transform have been proposed for the analysis of nonstationary data (we will limit the

* Received 22nd December, 1994.

** Faculty of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

*** Faculty of Engineering, Prince of Songkla University Hatyai, Songkla 90110, Thailand

definition of the term "nonstationary data" to the time series data with a spectrum which is time-varying when viewed using an adequately long time scale, but stationary when using a short time scale). Their applications can be found in fields such as speech processing. The methods are reported to have various advantages and disadvantages.

Thus, we feel that there must be a method to accurately obtain the time-variant spectrum of time series data such as those from the vibration of an engine with varying speeds, and the autoregressive (or AR) methods were selected due to their advantage in accurately estimating the spectrum even from a very small data set. In this study, the fundamental characteristics of the related methods will first be clarified, and clear guidelines for various applications will be developed. In this paper, time series generated from multiple sinusoidal signals were used for analysis and investigation. Non stationary data, as described above, will be discussed in a separate paper.

2. Objectives

In case of analyzing a short section of transient time series data which can be treated as stationary, the problem is to clarify the relationship between the degree of stationariness and the accuracy of the analysis. For instance, in dealing with a set of data with frequency changing at a constant rate, using a long data section will generally result in a good accuracy, but the characteristics of the first subsection and the last subsection will be very different. In contrast, using a short data section will result in good stationariness, but obtaining an accurate spectrum can be difficult. Direct investigation of the effects of various parameters on the analysis of nonstationary time series would be very complicated and might yield to misleading results. In this paper, we therefore limit our study to only stationary data to attain a clear understanding mainly of the relationship between the data length and other parameters.

The methods employed are AR methods. Investigation on the accuracy of the analysis and the applicability was carried out and the results were compared to those from the FFT method. To date, many AR methods have been proposed; three of these, which are shown in the appendix, were investigated. Various synthetic time series data were used, and the evaluation was done based on adequate trials. The data series generated was based on the vibration data of rotating machinery. The spectrum of this kind is usually line spectrum in which each line indicates a multiple of the rotational frequency. This is simply the combination of multiple sinusoidal signals with frequencies equal to multiples (and half order of the multiples) of a base frequency. Using data generated in this way, the accuracy performance of each approach was investigated.

All synthesized data were digitized; round-off error and other instrumentation noise were also taken into account by adding random noise. The S/N ratio of data series is 40 dB unless otherwise specified. To generalize the results, frequency was normalized as a fraction of sampling frequency and only values from 0.0 to 0.5 were used.

3. Spectral Analysis by AR Methods

3.1 Principle of AR methods

The principle of AR methods is to fit the observed data to an AR model, represented by

$$x[n]=-\sum_{k=1}^{p}a[k]x[n-k]+u[n-k], \qquad (1)$$

where p: model order, $a[k]$: AR coefficients, $u[n]$: input to AR model (usually white noise with variance σ_u^2), $x[n]$: observed data series.

Equation (1) is usually referred to as a p-order autoregressive model or simply AR(p) model.

Various approaches have been proposed to estimate the AR coefficients[5),6)]. In the first stage, autocorrelation method, covariance method, and Burg's method were adopted. The most effective one was then further investigated. The details of each method can be found in the appendix.

3.2 Spectrum determination

After AR coefficients are obtained using one of the methods presented in the appendix, the power spectral density (PSD) can be calculated from

$$P[f]=\frac{\Delta\sigma_u^2}{\left|1+\sum_{k=1}^{p}a[k]e^{-j2\pi f\Delta}\right|^2}. \qquad (2)$$

The PSD obtained using Eq.(2) is merely the PSD of the AR model fitted to observed data. It is visually appealing, but does not necessarily represent the real spectrum of the data. Moreover, the values and the area under each peak are unrelated to the amplitudes of the sinusoidal components in the data. The reason is that Eq.(2) was developed for obtaining the spectrum of the real AR process which is random in nature. The spectrum in this form is therefore not appropriate for estimating the spectrum of the sinusoidal signal, which is deterministic. In this study, the spectrum was estimated by utilizing the roots of the characteristic equation developed using the estimated AR coefficients. The procedure is as follows.

The frequencies and damping ratios of the components in the data are first determined from Eqs.(3) and (4), respectively.

$$f_i=\frac{1}{2\pi\Delta}\tan^{-1}\left(\frac{Im(z_i)}{Re(z_i)}\right) \qquad (3)$$

$$\zeta_i = \frac{\ln|z_i|}{\Delta} \qquad (4)$$

Here z_is are roots of the characteristic equation. These values are then used to form a complex damped sinusoidal model and fitted to the observed data, as shown in Eq.(5).

$$x[n] = \sum_{i=1}^{p} A[i] \cos(2\pi n f_i \Delta + \phi_i) e^{-\zeta_i n \Delta} \qquad (5)$$

This forms a set of linear simultaneous equations which can be solved for the unknown vector of complex amplitudes $A[i]$, and thus the phase angles ϕ_i. The spectrum is then obtained from the plot of these amplitudes versus the corresponding frequencies. In this paper, only undamped sinusoidal signals are considered; the damping ratios are equal to zero.

By this technique, the dependence of frequency resolution on the data length (or more precisely, time-scale) as in the case of the FFT method, is removed. This clearly shows the advantage of the AR approaches over the FFT approaches.

All computations are based on the programs in Ref.(6). The code was modified to do operations in double precision. The input sampled data series is a 16-bit integer array. The computer was an engineering workstation. Unless the model order is very high, very little CPU time was consumed.

4. Results and Discussion

4.1 Comparison of the performance of the 3 methods

To distinguish general characteristics of all 3 methods, a data series comprised of 3 sinusoids was generated. The frequencies and amplitudes were 0.2, 0.3, 0.4 and 10, 10, 8, respectively. First, a sufficiently long (2000) data record was used to examine the effects of model order on the resultant spectrum. The results are presented in Fig. 1. It is evident that the covariance method gives the best accuracy. Although it shows slight error in amplitude estimates for model order 6, but it gives very accurate spectrum estimates for orders 8 and higher. For data which consist of 3 harmonic components, model order 6 is adequate. Here orders higher than 6 must be used because the data are injected with random noise. The extra model orders are needed to model the noise components. Usually the contribution of these noise components is small for sufficiently high S/N ratio, and will not show up in the linear plot of the amplitude spectrum.

The autocorrelation method, on the other hand, gives both fault frequencies and amplitudes for model order 6. Increasing model order improves the frequency estimates, but still gives incorrect amplitude estimates. Moreover, extra fault components also show up with increasing order.

Burg's method is shown to accurately determine only the frequencies, but fails to estimate the amplitudes. Spurious peaks also occur with increasing model order.

Next, the effects of model order and data length on the spectrum estimates were examined. Fig. 2 shows a set of spectra with varying data length and the model order fixed at 18. It is clear that for the

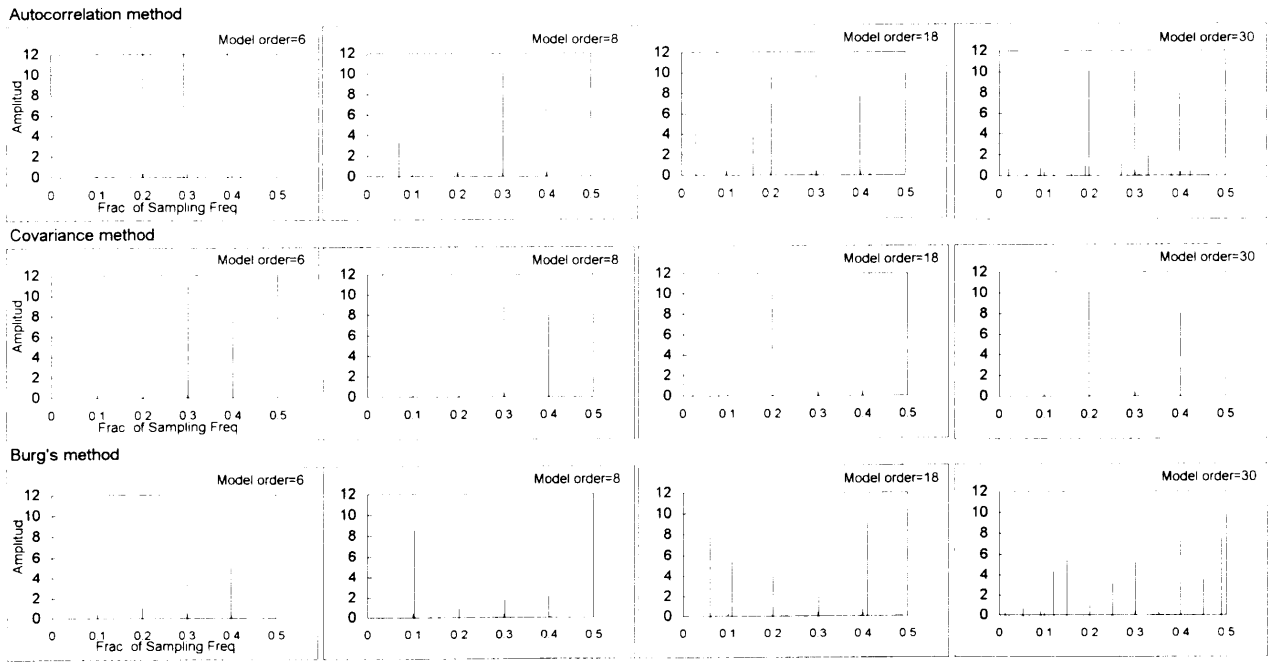

Fig. 1 Spectra diagrams for comparison of three AR methods

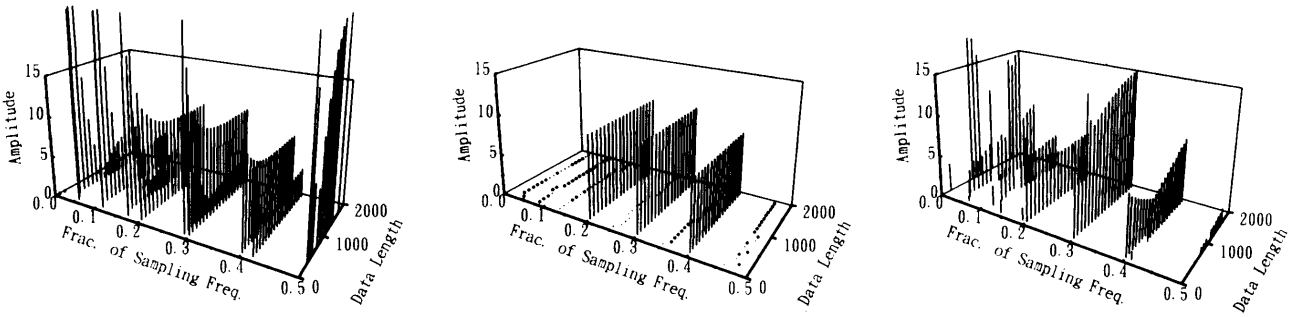

Fig. 2 Spectrum maps (Effect of data length)

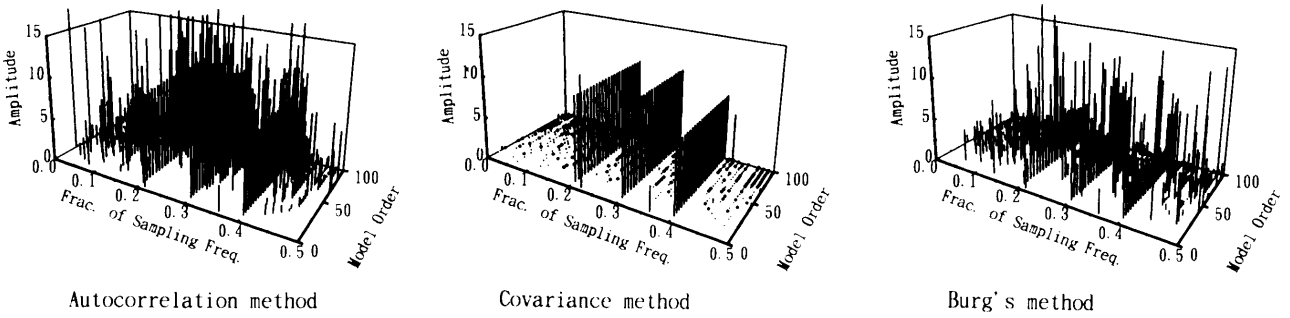

Fig. 3 Spectrum maps (Effect of model order)

covariance method, the spectrum estimates are not effected by the data length being used. The extra peaks, if they exist, are insignificant. For the autocorrelation method, though the amplitude estimates become better with increasing data length, the value of many fault peaks is relatively high and could not be identified if not known a priori. Burg's method shows no improvement for either frequency or amplitude estimates no matter how long the data record is. With a shorter data record, Burg's method seems to give better amplitude estimates. From the definition of each method, the differences shown here are not unexpected.

The effect of model order on the spectrum estimates was further investigated for better understanding of the performance of each method. This time a data length of 1000 was used, and the results are presented in Fig. 3. Burg's method and the autocorrelation method show many significantly high spurious peaks, and the amplitude estimates are also inaccurate. The covariance method, in contrast, shows accurate estimates of both frequencies and amplitudes regardless of the model order specified, as long as the order is greater than twice the number of frequency components. Although there are a number of methods for selecting the order for AR approaches, e.g. AIC and FPE, we found that these methods are suitable for fitting random time series. For a sinusoidal signal which is deterministic, these methods fail to show a minimum value at a specific model order. Fortunately, the covariance method is shown to be capable of yielding accurate spectrum estimates, even for data with an unknown number of frequency components, if high model order is specified. Theoretically, the model order must only be twice the number of frequency components in the data. The actual signal, however, is always contaminated by kinds of noise. Specification of theoretical model order generally does not eable the extraction of all the correct frequency components. The problem here is identifying how large a model order should be specified. The answer depends on the S/N ratio and other characteristics of the data. With many trials, we suggest that five times the number of frequency components will generally give adequate results. For data with frequency components very closely spaced, it will be shown later that a much larger model order is needed.

Fig. 4 shows the result from the same set of data with S/N ratio reduced to 20 dB. Since Burg's method and the autocorrelation method failed to yield good results, only the results from the covariance method are presented. If model order is specified much higher than necessary, the spectrum estimates exhibit spurious peaks. It is also evident that these spurious peaks

Table 1 Trial conditions (high-frequency range)

frequency	sample number per cycle	number of cycles L=512	number of cycles L=8
0.476	2.10	244	3.81
0.455	2.20	233	3.64
0.435	2.30	223	3.48
0.417	2.40	214	3.34
0.400	2.50	205	3.20
0.391	2.56	200	3.13
0.385	2.60	197	3.08
0.370	2.70	189	2.96

Table 2 Results of single frequency (high-frequency range)

	L=512				L=64			L=32		L=8	
Freq.	p=2	p=10	p=30	p=60	p=2	p=10	p=30	p=2	p=10	p=2	p=4
0.476	B	A	A	A	A-	A	A	B	A	A	B
0.455	B	A	A	F	A-	A	A-	A-	A	A	C
0.435	A-	A	A	A	A-	A	A	A-	A	A	C
0.417	A-	A	A	A	A	A	A	A	A	A	A
0.400	A-	A	A	A	A	A	A	A	A	A	A
0.391	A-	A	A	A-	A	A	A	A	A	A	A
0.385	A-	A	A	A	A	A	A	A	A	A	A
0.370	A-	A	A	A	A	A	A	A	A	A	A

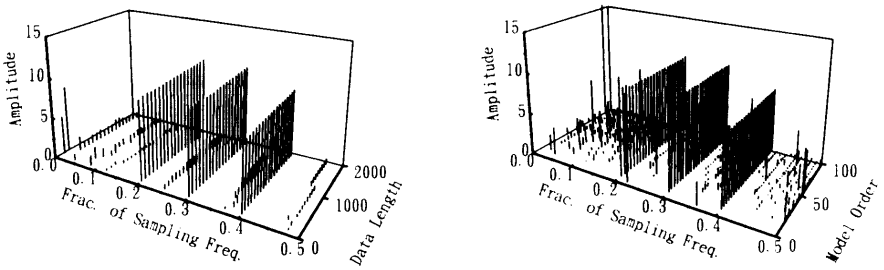

Fig. 4 Spectrum maps (Effect of S/N ratio)

do not occur at the same frequencies for different model orders, they can therefore easily be identified by comparing the results of data fitted the data using different model orders. The figure clearly indicates that the data length has no significant adverse effect on the accuracy of spectrum estimates. The covariance method is thus a robust method which is also capable of obtaining spectrum estimates of low S/N ratio time series with good accuracy.

4.2 Performance in estimating spectrum of extremely high and low frequencies —single sinusoidal signal—

In order to understand the accuracy performance of AR methods in estimating the spectrum of data which consists of extremely high and low frequencies (relative to sampling frequency), here the results using only one sinusoidal component will first be presented. For extremely high-frequency the number of data points per cycle is very small, as is the number of cycles in the data record for the case of extremely low-frequency.

Since the fraction of sample frequency was employed throughout this paper, the extremely high-frequency is thus the one with the fraction value very close to 0.5. A set of selected frequencies, the corresponding number of data points per cycle, and number of cycles per prescribed data length are presented in Table 1. To generalize the results obtained, the generated data series are adjusted to have a S/N ratio of 40 by injecting random noise. The accuracy performance of the covariance method with varying data length and model order was studied. Due to the large number of results, only tabulated results are given here. The performance rank, which is used for the estimation of results, is summarized as follows.

rank A : Excellent Results
rank A- : Fairly Good Results
rank B : Difference in Amplitude (5-20%)
rank C : Difference in Frequency (1-5%)
rank D : Both B and C
rank E : Large Difference in Frequency (5-10%)
rank F : Large Extra Spectral Lines
rank G : Lack of Some Spectral Lines
rank X : Absence of Correct Spectral Line

Table 2 presents the results for the case of extremely high-frequency. With data length of 512, which is more than sufficient, both too low and too high a model order reduces the performance of estimation. Since the data are a single sinusoid contaminated with noise, the theoretical model order of 2 is not sufficient. A model order of 10 yields excellent results. Increasing model order beyond this causes spurious peaks, and introduces slight error in the amplitude estimates, but this is insignificant. The results for the cases of data length of 64 and 32 are acceptable regardless of model order. For an extremely short data record with only 8 points of data, accurate spectrum estimates can be obtained with an order of 2 rather than 4. Moreover, with frequency as high as 0.495 times sampling frequency, it was observed that the spectrum could be accurately estimated using a data length of 32 and model order of 10. The results of this investigation indicate that unless the data length is insufficient, accurate spectrum estimates can

be determined using the covariance method. In the case of FFT, if data length is 32 points the frequency resolution is approximately 0.03. If the data number is not a multiple of cycle, side lobes will appear and accurate spectrum estimates cannot be obtained.

Next, results from data containing an extremely low-frequency component will be presented. In this case, the collected data will usually be less than one cycle. Since FFT assumes the available data length or time-width as its period, the correct result is unexpectable. The AR methods, on the other hand, do not possess such limitations. Frequency resolution can almost be arbitrarily chosen. Furthermore, it is reported that the AR methods are capable of estimating spectra of data with only half cycle. In this report, the data series summarized in Table 3 were used to examine the performance of the covariance method.

The results are presented in Table 4. In this extremely low-frequency case, the performance depends on the number of cycles contained in the data record, not the frequency itself. With data containing more than 0.4 times the component cycle, accurate spectrum estimates can be obtained without major difficulty. If this becomes 0.256, the estimation becomes unstable. Although no clear threshold number can be drawn from the trials, 0.4 times the cycle would be a good choice. Based on this number, the minimum frequency can be as low as 0.0005 (or 1/1000 of maximum frequency). This is another advantage of the AR method over FFT. It is also observed that for extremely low-frequency, higher order is needed. If, however, data contains 0.5 times the cycle, model order of 10 is sufficient.

Table 3 Trial conditions (low-frequency range)

frequency	number of cycles					(f=fc/n)
	L=512	L=128	L=64	L=32	L=8	n
0.0005	0.256	0.06	0.03	0.02	0.004	1000
0.001	0.512	0.13	0.06	0.03	0.008	500
0.002	1.024	0.26	0.13	0.06	0.016	250
0.004	2.048	0.51	0.26	0.13	0.032	125
0.010	5.12	1.28	0.64	0.32	0.08	50
0.013	6.4	1.6	0.8	0.4	0.1	40
0.025	12.8	3.2	1.6	0.8	0.2	20
0.050	25.6	6.4	3.2	1.6	0.4	10

Table 4 Results of single frequency (low-frequency range)

	L=512				L=128				L=64			L=32		L=8	
Freq.	p=2	p=10	p=30	p=60	p=2	p=10	p=30	p=60	p=2	p=10	p=30	p=2	p=10	p=2	p=4
0.0005	X	X	A-	A-	X	E	A	E	X	X	E	X	X	X	X
0.001	X	D	A	F	X	C	A	D	X	X	X	X	X	X	X
0.002	X	B	A	A	X	C	A	C	X	X	E	X	X	X	X
0.004	X	A-	A	A	X	A	A	A	X	C	C	E	E	X	X
0.010	D	A	A	A	D	A	A	A	D	A	A	D	C	X	X
0.0125	D	A	A	A-	D	A	A	A	D	A	A	D	A-	X	X
0.025	B	A	A	A	B	A	A	A	B	A	A	B	A	D	D
0.050	B	A	A	A	A-	A	A	A	A-	A	A	A	A	A-	D

We conclude here that the covariance method can be effectively used to estimate the spectrum for data which contains at least 0.4 times the cycle.

4.3 Performance for double sinusoidal components

Next, the accuracy performance of analyzing data containing two sinusoidal components was investigated. Two cases were considered here, i.e., components of two widely separated, and two adjacent frequencies.

(1) Data with two widely separated frequency components

This section presents the performance of analyzing data with two widely separated frequency components. The high-frequency was fixed at 0.45, and the low-frequency was varied from 1/20 to 1/1000 of the higher one. The results are presented in Table 5. It is seen that the method is capable of locating the correct spectrum of both components up to a ratio of only 1/100. A small deviation takes place in the case of a ratio of 1/200. The same trend was observed for shorter data records. With additional trials, the low-frequency component was seen to dominate the analyzing performance. As mentioned in the previous section, the fraction of the low-frequency component cycle directly affects the performance of analysis. The smaller the fraction, the larger the required model order. Generally, data containing frequencies separated by more than 1/100 can be analyzed by employing the techniques of filtering, there is thus little benefit to proceeding further to more extreme cases. Here we summarize that for data with the ratio of high and low frequencies up to 1 : 100 and the lower frequency component which contains at least 0.5 times its cycle, the corresponding spectrum can be determined accurately using the covariance method with properly specified model order.

(2) Data with two adjacent frequency components

For closely spaced frequency components, two base frequencies, i.e., 0.05 and 0.3, were considered. With each base frequency, another component with frequency equal to its base value plus a fraction,

Table 5 Results of double frequencies (widely separated)

f1=0.45,f2=f1/n		L=512			L=128
f2	n	p=8	p=20	p=40	p=40
0.00045	1000	G	G	C	G
0.0009	500	G	F	A-	C
0.00225	200	G	A-	A-	A
0.0045	100	B	A	A-	A
0.009	50	A-	A	A	A
0.0225	20	A	A	A-	A

Table 6 Results of double frequencies (closely spaced)

Diff. in two Freq.	n	f1=0.05,f2=f1+0.5/n L=512					f1=0.30,f2=f1+0.5/n L=512				
		p=8	p=20	p=40	p=60	p=80	p=8	p=20	p=40	p=60	p=80
0.0005	1000	G	G	B	F	F	G	G	B	D	A-
0.001	500	G	E	A-	A	A	G	D	B	A-	A
0.005	100	E	A	A	A	A	B	A	A	A-	A
0.01	50	D	A	A	A	A	B	A	A	A	A

Table 7 Results of multiple frequencies (3 lines)

f1=0.45,f2=0.2,f3=0.01

	L=							
	20	30	40	50	60	70	80	90
p=0.3L	G	A	A	A	A	A	A	F
p=0.4L	A-	A-	A-	A	A-	A	A	A

varied from 1/50 to 1/1000, of the Nyquist frequency (i.e., 0.5) was added to form a test data series. The results are as shown in Table 6. With frequency difference of about 1/500 of Nyquist frequency (or 0.001 of sampling frequency), the spectrum can be identified using a rather high model order. The same results were observed for both base frequencies. Model order of 80 is an extremely large value. The explanation for these results is that for identifying the spectrum of very closely spaced components, very high resolution is needed. Increasing model order is thus equivalent to increasing the frequency resolution. If the frequency difference is only 0.005, a model order of 20 is sufficient. In any case, if the FFT method were used, the resolution would be 0.002 for a data record of 512 points. This clearly limits the ability to separate 2 adjacent frequency components. This shows that this AR method is superior to FFT. Supplemental trials confirm the effectiveness of the method as long as data length for low frequency is adequate.

4.4 Data with multiple frequency components

This section presents the performance of multiple components analysis, i.e., 3, 10, and 20 components with a pair of adjacent frequency components. Table 7 shows the results for data with 3 frequency components. It is observed that, with the exception of a very short data record, the spectrum can be obtained without any difficulty. The data length, as mentioned in the previous section, depends on the component with lowest frequency. Likewise, if a pair of adjacent frequency components exists, high model order is needed.

The details for data containing 10 and 20 components are shown in Table 8. The analyzing performance is summarized in Table 9. Results for short and sufficiently long data records are presented for

Table 8 Trial conditions (multiple frequencies)

cond.	frequencies included
10 lines	0.01, 0.05, 0.10, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.47
20 lines (1)	0.01, 0.03, 0.05, 0.06, 0.08, 0.09, 0.10, 0.14, 0.16, 0.18, 0.20, 0.22, 0.26, 0.30, 0.35, 0.38, 0.40, 0.44, 0.46, 0.47
20 lines (2)	0.01, 0.03, 0.05, 0.06, 0.08, 0.09, 0.10, 0.14, 0.16, 0.18, 0.20, 0.22, 0.26, 0.30, 0.35, 0.395, 0.40, 0.44, 0.46, 0.47
20 lines (3)	0.01, 0.03, 0.0495, 0.06, 0.08, 0.09, 0.10, 0.14, 0.16, 0.18, 0.20, 0.22, 0.26, 0.30, 0.35, 0.38, 0.40, 0.44, 0.46, 0.47

Table 9 Results of multiple frequencies (10 and 20 lines)

10 Lines	L=							
	50	60	70	80	90	100	110	120
p=24	A-	A-	A	A	A	A	A	A

20 Lines	L=							
	130	150	180	200	240	280	300	320
(1)p=44	G	G	G	G	G	G	G	G
50	B	B	B	B	B	B	B	B
60	A-	A	A	A	A	A	A	A
(2)p=60	B	A-	A	A	A	A	A	A
(3)p=60	G	G	A-	A	A	A-	A-	A-
80	G	B	A	A	A	A	A-	A-

both cases. For data with 10 components, an accurate spectrum can be obtained with short data record and relatively low model order. For data with 20 components, on the other hand, relatively high model order is needed to resolve the spectrum, and at least 180 data points are required. As can be expected, much higher model order is needed if adjacent frequency components exist. Supplemented trials have led to the conclusion that in the case of a data length of about 10 times the number of components, as long as it is not less than 0.4 times the cycle of the lowest frequency component, a sufficiently accurate spectrum can be resolved using the covariance method.

5. Conclusions

In this paper, investigation results on the accuracy performance of spectrum analysis for multiple sinusoidal data using AR methods are presented. The following conclusions can be drawn.

(1) The PSD obtained from the AR coefficients using Eq.(2) provides no useful information for extracting the real amplitude of the frequency components. The frequencies and damping ratios must first be determined from the roots of the characteristic equation, and amplitudes and phases then obtained by fitting a damped sinusoidal model to the data.

(2) Various methods were proposed to obtain the AR coefficients; the covariance method has been confirmed to be the most effective one with the results independent of data length and model order. The lack of dependence of performance on model order is a very important characteristic which has eliminated

the cumbersome procedure of selecting a proper model order.

(3) The results of investigation using synthesized data have led to the understanding of the limitations and characteristics of the covariance method. For an extremely high-frequency component (relative to Nyquist frequency), the correct spectrum can be estimated without restriction even with a short data record. For extremely low-frequency, on the other hand, the spectrum analysis becomes impossible if the data record amounts to less than 0.4 times its cycle.

(4) If the data contain a sufficient number of points for the low-frequency component as stated above, the spectrum can be accurately estimated for both widely separated and very closely spaced frequency components by using the covariance method.

(5) If not contradicted with the conditions mentioned in (3), the guideline for assigning the data length is to use at least about 10 times the number of frequency components embedded in the data. A good choice for the model order is about half of this number.

(6) Since the above characteristics are not realizable by the FFT method, this thus reveals the advantage of AR methods in application to spectrum analysis.

Appendix

(1) Autocorrelation method

Multiplying both sides of Eq.(1) and then taking the expected value, with the autocorrelation function defined as in Eq.(6) and taking into account the properties of $u[n]$, will result in Eq.(7).

$$r[k]=\begin{cases}\frac{1}{N}\sum_{n=0}^{N-1-k}x[n]x[n+k] & (k:\ 0\sim p)\\ r[-k] & (k:\ -1\sim-(p-1))\end{cases} \tag{6}$$

$$r[k]+\sum_{n=1}^{p}a[n]r[k-n]=0 \tag{7}$$

Rewriting the above equation in the matrix form yields the Yule-Walker equation

$$\begin{bmatrix} r[0] & r[1] & \cdots & r[p-1]\\ \vdots & \ddots & \cdots & \vdots\\ r[p-1] & \cdots & \cdots & r[0]\end{bmatrix}\begin{bmatrix}a[1]\\ \vdots\\ a[p]\end{bmatrix}=-\begin{bmatrix}r[1]\\ \vdots\\ r[p]\end{bmatrix} \tag{8}$$

which can be efficiently solved for $a[1], \cdots, a[p]$ using Levinson's algorithm. Since this method assumes 0 outside the observed data, its performance is acceptable for long data records, but its resolution is reduced for short data records.

(2) Covariance method

If the correlation function is taken in the form of Eq.(9) instead of Eq.(6), the matrix form of Eq.(7) becomes as shown by Eq.(10).

$$c[i,j]=\frac{1}{N-p}\sum_{n=p}^{n-1}x[n-i]x[n-j] \tag{9}$$

$$\begin{bmatrix} c[1,1] & c[1,2] & \cdots & c[1,p]\\ c[2,1] & c[2,2] & \cdots & \vdots\\ \vdots & \vdots & \ddots & \vdots\\ c[p,1] & \cdots & \cdots & c[p,p]\end{bmatrix}\begin{bmatrix}a[1]\\ \vdots\\ \vdots\\ a[p]\end{bmatrix}=-\begin{bmatrix}c[1,0]\\ c[2,0]\\ \vdots\\ c[p,0]\end{bmatrix} \tag{10}$$

Equation (10) is different from Eq.(8) in that the covariance matrix is not a Toeplitz one, and thus Eq. (10) cannot be solved using Levinson's algorithm. This matrix is, fortunately, Hermatian and the efficient Cholesky decomposition can be employed to determine AR coefficients.

An alternative approach to derive the covariance method is to apply the least squares method directly to Eq.(1) which is just another form of Prony's method.

(3) Burg's method

In contrast to the above methods which estimate AR coefficients directly, another method proposed by Burg first estimates the reflection coefficients and then uses Levinson recursion to calculate the AR coefficients. The computation procedures are as follows.

The forward and backward prediction errors are first obtained from

$$\hat{\rho}_k^f=\frac{1}{N-k}\sum_{n=k}^{N-1}\left\|x[n]+\sum_{i=1}^{k}a_k[i]x[n-i]\right\|^2 \tag{11}$$

$$\hat{\rho}_k^b=\frac{1}{N-k}\sum_{n=0}^{N-1-k}\left\|x[n]+\sum_{i=1}^{k}a_k[i]x[n+i]\right\|^2, \tag{12}$$

where

$$\hat{a}_k[i]=\begin{cases}\hat{a}_{k-1}[i]+k_k a_{k-1}[\hat{k}-i] & (i=1,2,\cdots,k-1)\\ k_k & (i=k).\end{cases} \tag{13}$$

If the $(k-1)$th reflection coefficient is known, the prediction error (average of forward and backward prediction errors) is thus only a function of $k(k)$. Minimizing prediction error yields the estimates of $k(k)$:

$$k_k=\frac{-2\sum_{n=k}^{N-1}e_{k-1}^f[n]e_{k-1}^b[n-1]}{\sum_{n=k}^{N-1}(\|e_{k-1}^f[n]\|^2+\|e_{k-1}^b[n-1]\|^2)}, \tag{14}$$

where

$$e_k^f[n]=x[n]+\sum_{i=1}^{k}a_k[i]x[n-i] \tag{15}$$

$$e_k^b[n]=x[n-k]+\sum_{i=1}^{k}a_k[i]x[n-k+i]. \tag{16}$$

Substituting Eq.(13) into Eqs.(15) and (16) yields the following recursive expressions:

$$e_k^f[n]=e_{k-1}^f[n]+k_k e_{k-1}^b[n-1] \tag{17}$$

$$e_k^b[n]=e_{k-1}^b[n-1]+k_k e_{k-1}^f[n], \tag{18}$$

where

$$e_0^f[n]=e_0^b[n]=x[n]. \tag{19}$$

After the kth reflection coefficient is obtained, the corresponding AR coefficients can be readily calculated from Eq.(13). Burg's method is known to

produce more accurate AR spectrum estimates compared to correlation-based methods. Shortcomings, e.g., frequency bias as a function of initial phases and spectral line splitting, were also reported.

References

(1) Kesler, S. B., Ed., Modern Spectrum Analysis II, (1986), IEEE Press.

(2) Suzuki, K., Studies on Identification of Structural Dynamics (Survey of Current Works and Proposal of a New Approach), Trans. Jpn. Soc. Mech. Eng., (in Japanese), Vol. 54, No. 501, C (1986), p. 1035.

(3) Pandit, S. M. and Wu, S. M., Time Series and Systems with Applications, (1990), Krieger.

(4) Pandit, S. M., Modal and Spectrum Analysis: Data Dependent Systems in State Space, (1991), Wiley-Interscience.

(5) Kay, S. M., Modern Spectral Estimation: Theory and Application, (1988), Prentice-Hall.

(6) Marple, S. L., Digital Spectral Analysis with Applications, (1987), Prentice-Hall.