A CLASS OF MODELS FOR IDENTIFICATION AND SIMULATION OF EARTHQUAKE GROUND MOTIONS

R. M. OLIVER *, K. S. PISTER **
University of California, Berkeley, California 94720, U.S.A.
** Department of Industrial Engineering & Operations Research*
*** Department of Civil Engineering*

SUMMARY

This paper outlines the use of discrete, autoregressive/moving-average (ARMA) models for identification and estimation of parameters in models derived from analysis of uniformly digitized earthquake ground motion acceleration data. Such models are of equal generality as compared to continuous-time models and have a number of significant advantages for purposes of digital analysis and simulation. The structure of ARMA models is briefly described, their relation to continuous models noted, and results of their application to a number of recorded accelerograms summarized.

1. Introduction

Most existing models for the analysis and simulation of earthquake ground motion records are formulated in continuous time, using linear differential equations with inhomogeneous forcing functions given by white noise which, in certain cases, has been assumed to be correlated. Typically, the order of the linear differential equations and the degree of correlation in the noisy forcing function is specified for theoretical or practical reasons. The coefficients of these differential equations are expressed in terms of the natural frequencies and damping constants of second-order harmonic oscillators, where appropriate values for these parameters are usually obtained by matching certain predominant spectral characteristics of real earthquake records with those obtained from the differential equation models. Simulated accelerograms are then generated digitally by numerical integration of the differential equation or impulse response function (with white noise input), or else by using the theoretical Fourier amplitude spectrum (based on the transfer function) to weight a superposition of a large number of sinusoids at equispaced frequencies with randomly generated phase angles. The white noise input or filtered noise output is generally multiplied by an appropriate envelope function to incorporate non-stationary characteristics (i.e., build-up and decay). In view of the current availability of large quantities of uniformly digitized earthquake acceleration data for analysis, and the widespread interest in generating artificial digitized accelerograms for structural response studies, it appeared worthwhile to consider the use of models formulated explicitly in discrete time. Autoregressive/moving-average (ARMA) models are an important class of discrete models which can be represented as stochastic linear difference equations of finite order. ARMA models are of equal generality with linear continuous-time models (differential equations), but they have a number of significant advantages for purposes of digital analysis and simulation. A large body of literature, exemplified by the work of Box and Jenkins [1], gives systematic procedures for identifying the order of the ARMA model which best describes a particular time series (such as a digitized accelerogram) based on time-domain analysis of the actual data (i.e., without a priori assumptions). Moreover, maximum-likelihood techniques are available for estimating optimal parameter values directly from the data, with specifiable confidence intervals for the estimates. The sequence of residuals--i.e., deviations from the fitted model, or "one-step-ahead forecast errors"-- provides a basis for quantitative statistical tests of goodness of fit, and represents a direct estimate of the underlying noise sequence driving the observed process. These time-domain analytic techniques are somewhat less sensitive than frequency-domain techniques (e.g., spectral analysis) to certain violations of stationarity assumptions and to the effects of digitizing a continuous record. ARMA models can be used directly for discrete simulation by simple iteration of the difference equations, with appropriate discrete noise input, thus simplifying the procedure of obtaining artificial accelerograms with characteristics similar to specified real accelerograms.

In this study, the ARMA model-identification and parameter-estimation techniques of Box and Jenkins were applied to a number of California earthquake records, and the results suggest that two particular ARMA models are worth considering in some detail. These are the second-order-autoregressive/first-order-moving-average (ARMA(2,1)) model and the fourth-order-autoregressive/first-order-moving-average (ARMA(4,1)) model, which may be considered to correspond to continuous-time models described, respectively, by second- and fourth-order differential equations.

2. The ARMA (2,1) Model

The ARMA (2,1) model for a stationary correlated process a_t is defined by the 2nd-order-autoregressive/1st-order-moving-average difference equation:

$$a_t - \phi_1 a_{t-1} - \phi_2 a_{t-2} = e_t - \theta_1 e_{t-1} , \tag{1}$$

in which it is assumed that $e_t \sim N(0,\sigma_e^2)$ is independently and identically distributed. That is to say, the input is stationary discrete white noise. In terms of the backward shift operator B (defined by $B^k x_t = x_{t-k}$) eq. (1) can be rewritten:

$$\left(1 - \phi_1 B - \phi_2 B^2\right) a_t = (1 - \theta_1 B) e_t , \tag{2}$$

or equivalently in the factored form:

$$(1 - r_1 B)(1 - r_2 B) a_t = (1 - \theta_1 B) e_t , \tag{3}$$

where r_1 and r_2 are the roots of the characteristic polynomial,

$$r^2 - \phi_1 r - \phi_2 = 0 . \tag{4}$$

A requirement for stationarity (stability) of the process a_t is that the autoregressive roots r_1 and r_2 lie within the unit circle, or equivalently, that $|\phi_2| < 1$, $\phi_1 + \phi_2 < 1$ and $\phi_2 - \phi_1 < 1$.

The autocorrelation function of the process a_t is symmetric in lag k so that

$$\rho_k = \text{cov}\,[a_t, a_{t+k}]$$

$$= \sigma_a^{-2}\ \text{cov}\,[a_t, a_{t+k}] = \rho_{-k} \qquad k = 1,2,\ \ldots \tag{5}$$

where the variance of the output process a_t ,

$$\sigma_a^2 = \text{var}\,[a_t] = \frac{1 - \phi_2}{1 + \phi_2} \cdot \frac{\left(1 + \theta_1^2\right)\sigma_e^2}{(1 - \phi_2)^2 - \phi_1^2}\,, \tag{6}$$

is proportional to the variance σ_e^2 of the random forcing function e_t . It is well known that for $k \geq 2$ the autocorrelation function ρ_k must satisfy the homogeneous difference equation of eq. (1) or,

$$\rho_k - \phi_1\rho_{k-1} - \phi_2\rho_{k-1} = 0 \qquad k \geq 2 \tag{7}$$

with initial values given by

$$\rho_0 = 1$$

$$\rho_1 = \frac{\phi_1\left(1 + \theta_1^2 - \theta_1\phi_1\right) - \theta_1\left(1 - \phi_2^2\right)}{(1 - \phi_2)\left(1 + \theta_1^2 - \theta_1\phi_1\right) - \theta_1\phi_1(1 + \phi_2)}\,. \tag{8}$$

We note that ρ_1 depends on the moving average parameter θ_1 but that for $k \geq 2$ eq. (7) does not explicitly include θ_1 . In the ARMA (2,1) model of eq. (7) we are dealing with a system of 2nd-order linear difference equations whose solutions can be written as

$$\rho_k = c_1r_1^k + c_2r_2^k \qquad r_1\ ,\ r_2\ \text{ distinct} \tag{9}$$

where c_1, c_2 are derived from the initial values in eq. (8). When $\phi_1^2 < -4\phi_2$ the characteristic roots of eq. (4), r_1 and r_2 , are complex conjugates. The autocorrelation function can then be written in the form

$$\rho_k = (-\phi_2)^{k/2}\,\frac{\cos\,(k\lambda_d - \mu_d)}{\cos\,(-\mu_d)}\ ,\ k \geq 0 \tag{10}$$

where

$$\lambda_d = \cos^{-1}\left(\frac{\phi_1}{2\sqrt{-\phi_2}}\right) \tag{11}$$

has the interpretation of a frequency and

$$\mu_d = \tan^{-1}\left(\frac{2\rho_1 - \phi_1}{\sqrt{-\phi_1^2 - 4\phi_2}}\right) \tag{12}$$

has the interpretation of a phase angle. Since the autocorrelation of lag 1 depends on both autoregressive and moving average parameters in eq. (8), it follows that only the phase μ_d depends on the moving average parameter. One of the useful results of this study derives from eqs. (10), (11), (12) and the close analogies that the discrete model frequency and phase have with their continuous differential equation counterparts. Up to this point, the discrete model does not include an explicit time dimension for the lags $k = \pm 1, \pm 2, \ldots$, etc. If we let $\tau = k\Delta t$, i.e., each lag is separated by a time interval of length Δt then eq. (10) can be rewritten in the form

$$\begin{aligned} \rho(\tau) &= e^{\log(-\phi_2)\frac{\tau}{2\Delta t}} \frac{\cos\left(\frac{\lambda_d}{\Delta t}\tau - \mu_d\right)}{\cos\,(-\mu_d)} \\ &= e^{-\xi\omega_o\tau} \frac{\cos\left(\left(\omega_0\sqrt{1-\xi^2}\right)\tau - \mu_d\right)}{\cos\,(-\mu_d)} \quad \text{for} \quad \tau = \Delta t, 2\Delta t, \ldots \end{aligned} \tag{13}$$

where ω_0 has the interpretation of a natural frequency, and ξ that of a damping ratio.

$\frac{\lambda_d}{\Delta t} = \omega_0\sqrt{1 - \xi^2}$ may be thought of as the resonant frequency with ω_0 given by

$$\omega_0 = \frac{1}{\Delta t}\sqrt{\frac{1}{4}\,(\log\,(-\phi_2))^2 + \lambda_d^2} \tag{14}$$

and ξ by

$$\xi = \frac{-\log\,(-\phi_2)}{2\sqrt{\frac{1}{4}\,(\log\,(-\phi_2))^2 + \lambda_d^2}}\,. \tag{15}$$

We again note that only μ_d, the phase, depends on the moving average parameter, θ_1 ; the natural frequency ω_0 and the damping constant ξ depend only on the autoregressive coefficients. It is clear from equations (11), (12), (14), (15) that one can obtain the frequency ω_0 , damping constant ξ, and phase μ_d directly from estimates of the autoregressive and moving average parameters ϕ_1 , ϕ_2 , θ_1 of the discrete models in eq. (1) and eq. (7) without any need to estimate the spectrum of the underlying process.

3. Relations Between Discrete and Continuous 2nd-Order Models

Many earthquake acceleration models in the literature have been based on 2nd-order linear filters. The properties of these models can be put in correspondence with discrete ARMA models by utilizing the well-known correspondence between the statistical characteristics of the discrete (sampled) process a_t and the underlying continuous process $a(t)$. In particular, the (discrete) autocorrelation function (acf) of a_t, defined as $\rho_k = \text{cor}\,[a_t, a_{t+k}]$, and the (continuous) acf of a(t), defined as $\rho(\tau) = \text{cor}\ [a(t), a(t + \tau)]$, are equal at all points where both are defined, i.e., at integral multiples of the sampling interval. That is,

$$\rho_k = \rho(k\Delta t) \ , \ k = 0,1,2, \ldots \tag{16}$$

where Δt is the sampling interval, and $\rho_0 \equiv \rho(0) \equiv 1$. If a(t) is "low-pass-filtered" prior to sampling to eliminate power at frequencies greater than half the sampling frequency (i.e., at all $\omega > \frac{\pi}{\Delta t}$) -- or if a(t) initially contains negligible power at these frequencies-- then the power spectral density functions (psdf's) of a_t and a(t), which are the Fourier transforms of the corresponding acf's, will approximately coincide for frequencies in the range $0 \leq |\omega| \leq \frac{\pi}{\Delta t}$.

In general, a continuous random process described by an n^{th} order differential equation, when sampled at regular intervals Δt, gives rise to a discrete time series which is exactly described as an ARMA (n,n - 1) process. Various formulas can be used to obtain approximate conversion relationships between the parameters of the differential equation and the parameters of the corresponding ARMA model -- e.g., the differential operator, d/dt, can be approximated by a rational function of the backshift operator B, such as the backward differences, $(1 - B)/\Delta t$, or the trapezoidal formula, $2(1 - B)/(1 + B)\Delta t$. However, in the second-order case the exact conversion relationships can be readily obtained by enforcing eq. (16) and utilizing the results given in Section 2. In particular, if the sampling frequency is at least twice the resonant frequency $\left(\text{i.e.,}\ \frac{\pi}{\Delta t} \geq \text{Real}\left\{\omega_0\sqrt{1 - \xi^2}\right\}\right)$, then the frequencies and damping factors of the discrete and continuous acf's can be equated separately in eq. (16) to yield the following one-to-one conversion relationships between the autoregressive parameters (ϕ_1, ϕ_2) of the discrete process and the parameters (ω_0, ξ) of the continuous process (see [2] for details):

$$\phi_2 = -\exp\ (-2\omega_0 \xi \Delta t) \tag{17}$$

$$\phi_1 = 2\ \exp\ (-\omega_0 \xi \Delta t)\ \cos\left(\omega_0\sqrt{1 - \xi^2}\Delta t\right) \quad \text{if} \quad \xi \leq 1 \tag{18}$$

$$\phi_1 = 2\ \exp\ (-\omega_0 \xi \Delta t)\ \cosh\left(\omega_0\sqrt{\xi^2 - 1}\Delta t\right) \quad \text{if} \quad \xi \geq 1 \tag{19}$$

In the continuous-to-discrete conversion, after determining ϕ_1 and ϕ_2 from ω_0 and ξ, θ_1 can be determined from ϕ_1 , ϕ_2 and ρ_1 (where $\rho_1 = \rho(\Delta t)$) by solving:

$$\theta_1^2 + \left[\frac{2\rho_1\phi_1 - \phi_1^2 + \phi_2^2 - 1}{\phi_1 - \rho_1(1 - \phi_2)}\right]\theta_1 + 1 = 0 \qquad |\theta_1| \leq 1 \ . \tag{20}$$

Equation (20) is equivalent to eq. (8) expressing the autocorrelation of lag 1 in terms of model parameters.

4. The ARMA (4,1) Process and Its Interpretation

The ARMA (4,1) process is defined by the 4th-order-autoregressive/1st-order-moving-average linear difference equation:

$$a_t - \phi_1 a_{t-1} - \phi_2 a_{t-2} - \phi_3 a_{t-3} - \phi_4 a_{t-4} = e_t - \theta_1 e_{t-1} \tag{21}$$

in which $e_t \sim N(0,\sigma_e^2)$ as before. In terms of the backward shift operator this can be rewritten in factored form as:

$$(1 - r_1 B)(1 - r_2 B)(1 - r_3 B)(1 - r_4 B)a_t = (1 - \theta_1 B)e_t \tag{22}$$

where r_1, r_2, r_3, and r_4 are the roots of the characteristic polynomial

$$r^4 - \phi_1 r^3 - \phi_2 r^2 - \phi_3 r - \phi_4 = 0 \,. \tag{23}$$

If a time series is identified as an ARMA (4,1) process and the estimates of the autoregressive parameters are such that the characteristic polynomial has at least one complex pair of roots, then a unique factorization of the 4th-order autoregressive factor into two 2nd-order factors can be performed [2]. The result can be expressed:

$$\left(1 - \phi_{11}B - \phi_{12}B^2\right)\left(1 - \phi_{21}B - \phi_{22}B^2\right)a_t = (1 - \theta_1 B)e_t \tag{24}$$

where $\phi_{11} = (r_1 + r_2)$, $\phi_{12} = (-r_1 r_2)$, $\phi_{21} = (r_3 + r_4)$, and $\phi_{22} = (-r_3 r_4)$.

Physically, the ARMA (4,1) process represented in eq. (24) can be considered to arise from the action of an ARMA (2,1) filter and an AR (2) filter in series, as follows: white noise e_t first passes through an ARMA (2,1) filter to produce an intermediary process which serves as the input to an AR (2) filter whose output is a_t. The use of 2nd-order filters in series (in a different context) has been reported by Murakami and Penzien [3], among others.

5. Analysis of Earthquake Ground Motion Records

The data consisted of California Institute of Technology corrected accelerograms digitized at .02 seconds. Four earthquake records, comprising a total of six components were studied. Identification and parameter estimation for the ARMA models was performed using the systematic procedures of Box and Jenkins [1] employing the TIMES program documented in [4]. Forty seconds of record were examined, utilizing five-second (250 data point) windows. With one exception, all segments were fitted by ARMA (4,1) models, for which a set of "average" ARMA parameter values could be determined such that they would fall well within the 95% confidence intervals for the parameter estimates of nearly every segment. A detailed description of the data analysis including identification and estimation techniques, discussion of ARMA (4,1) and (2,1) models, and time-variation of parameter estimates within analysis windows can be found in [2]. Here space limits us to stating only the conclusions reached.

6. Conclusions

The application of the time-domain analytic techniques of Box and Jenkins to segments of digitized earthquake accelerograms appears to be a potentially useful method of characterizing recorded earthquakes by linear models with a small number of parameters. It should be emphasized that the discussions of the ARMA (2,1) and ARMA (4,1) models in this paper were entirely motivated by the experimental results -- no a priori assumptions were made concerning the order of appropriate ARMA models for earthquake analysis. The fact that the Box-Jenkins method includes systematic model-identification techniques which do not require such assumptions is one of its principal advantages relative to other model-fitting procedures commonly applied to earthquake data. Therefore, the applications of Box-Jenkins techniques to other earthquake records should not necessarily be expected to yield only ARMA (2,1) and ARMA (4,1) models. However, the fact that five of the six components studied here were best fitted in all their segments by an ARMA (4,1) model suggests that it may be appropriate for California earthquakes. Moreover, these models have appealing connections with simple hypothetical physical models discussed elsewhere in the literature. The ARMA (2,1) model may be considered to include the various basic forms of the linear-oscillator model; and the ARMA (4,1) model, in its representation as two linear oscillators with different natural frequencies acting in series, is somewhat more complex. It appears that the principal source of nonstationarity of the earthquake acceleration data lies in the time-dependence of variance of the driving noise process, rather than in the filtering parameters.

It may be significant that the one component which was fitted by the simpler ARMA (2,1) models -- the El Centro 1940 component -- was also the oldest record studied, suggesting that this difference might be in part due to differences in the way the data may have been recorded or processed. Also, it should be noted that all the data used in this study had been "corrected" in various ways, including low-pass filtering to prevent aliasing in spectral analysis. Since the Box-Jenkins techniques emphasize time-domain analysis, this low-pass filtering was not necessary (although other aspects of the correction process, such as the equal spacing of data points, were essential); therefore, it is possible that to some extent the models identified here reflect the corrective filtering rather than the physical process itself.

References

1. Box, G. E. P., and Jenkins, G. M., TIMES SERIES ANALYSIS: FORECASTING AND CONTROL, Revised Edition, Holden-Day, San Francisco, (1976).

2. Chang, M. K., Kwiatkowski, J. W., Nau, R. F., Oliver, R. M., and Pister, K. S., "ARMA Models for Earthquake Ground Motions," Report No. ORC 79-1, Operations Research Center, University of California, Berkeley, (Jan. 1979).

3. Murakami, M., and Penzien, J., "Nonlinear Response Spectra for Probabilistic Seismic Design and Damage Assessment of Reinforced Concrete Structures," Report No. EERC 75-38, Earthquake Engineering Research Center, University of California, Berkeley, (1975).

4. Willie, R. R., "Everyman's Guide to TIMES," Report No. ORC 77-2, Operations Research Center, University of California, Berkeley, (1977).