Nonstationary ARMA modeling of seismic motions

J.P. Conte

Civil Engineering Department, Rice University, Houston, Texas 77251-1892, USA

K.S. Pister

University of California, Santa Cruz, California 95064, USA

&

S.A. Mahin

Department of Civil Engineering, University of California, Berkeley, California 94720, USA

Discrete time-varying autoregressive — moving average (ARMA) models are used to describe realistic earthquake ground motion time histories. Both amplitude and frequency nonstationarities are incorporated in the model. An iterative Kalman filtering scheme is introduced to identify the time-varying parameters of an ARMA model from an actual earthquake record. Several model verification tests are performed on the identified model. Applications of these identification and verification procedures are given and show that the proposed models and identification algorithms are able to capture accurately the nonstationary features of real earthquake accelerograms, especially the time-variation of the frequency content. The well-known Kanai-Tajimi earthquake model is covariance equivalent with a subset of the low order ARMA(2,1) model. Using the results and methodology of this study, the parameters of a time-varying Kanai-Tajimi earthquake model can be estimated from a target earthquake record or they can be directly associated with characteristic earthquake features such as predominant frequency and frequency bandwidth.

Key words: earthquake ground motion, nonstationary, time-varying ARMA model, identification, verification, Kalman filtering.

INTRODUCTION

When designing earthquake-resistant structures, the ultimate goal of the earthquake engineer is to save human lives and reduce the extent of structural damage. However, the challenge consists in providing structural safety under economic constraints, which dictates a tolerance for some degree of structural damage inflicted during severe earthquake motions. If one accepts that part of the seismic input energy in the structure is dissipated by means of ductile inelastic deformations, the seismic loads to which the structure is subjected are

reduced (from the elastic condition). Presently, most seismic design codes rely implicitly or explicitly on ductile inelastic structural behavior to withstand severe earthquake motions. The task of the earthquake engineer is further complicated by two important facts: (i) the considerable uncertainty associated with the expected seismic excitation, and (ii) the possibility of high sensitivity of inelastic structural behavior with respect to the details of the ground motion time history.[3,14]

Considerable uncertainty is attached to each level of the ground motion generation mechanism. The precise nature of the fault rupture process, the characteristics of the wave propagation path and the local site conditions are all very uncertain. The resulting uncertainty

contained in the ground motion at a given site must be quantified and used in analysis, design, and risk assessment of structures. A probabilistic approach offers an attractive way to handle this complicated problem. For structural response prediction purposes, earthquake ground motions are commonly defined in terms of global parameters, (e.g. magnitude, epicentral distance, duration, peak ground motions, predominant frequency), or ground motion time histories consistent with the design site conditions. For example, when using the response spectrum analysis technique[9] to estimate the peak response of linear and nonlinear structures, the earthquake excitation is defined in terms of a design response spectrum which is obtained empirically from peak ground acceleration, velocity and displacement.[22,23] The uncertainty associated with the global ground motion parameters and the uncertainty carried by the detailed structure of the ground motion time history can be treated separately in a probabilistic conditioning setup. The first type of uncertainty can be handled using probabilistic seismic hazard analysis techniques.[4] To model the second type of uncertainty, the paradigm of viewing an actual earthquake ground motion as a single realization of an underlying stochastic earthquake process is used. Although stochastic earthquake models have been widely accepted in the engineering profession, many of them are not able to capture well enough all the features of actual seismic records which are important for structural response predictions.

The study reported in this paper deals only with the uncertainty contained in the ground motion time history and uses time-varying, linear, discrete, stochastic processes for realistic modeling of earthquake ground motions. The emphasis is placed on model identification, model verification, relationship with a classical continuous model, and investigation of the statistical distribution of traditional ground motion parameters using the fitted earthquake models.

ARMA MODELS FOR EARTHQUAKE GROUND MOTIONS

Stationary case

The general time-invariant ARMA model of order (p,q), abbreviated ARMA(p,q), is represented by the following stochastic linear difference equation:

$$a_k - \phi_1 a_{k-1} - \cdots - \phi_p a_{k-p} = e_k - \theta_1 e_{k-1} - \cdots - \theta_q e_{k-q}$$

$$(1)$$

In this equation, $\{a_k = a(k\Delta t),\ k = 0, 1, 2, \ldots\}$ represents the discrete earthquake ground acceleration process, Δt the sampling time interval, $\{e_k\}$ a zero-mean Gaussian white-noise of variance σ_e^2, and $(\phi_i,$

$i = 1, \ldots p)$ and $(\theta_i,\ i = 1, \ldots, q)$ the autoregressive and moving average coefficients, respectively. By using eqn (1) recursively, the ARMA(p,q) model can be expressed as a pure MA model of infinite order or ARMA$(0,\infty)$ if the stability conditions are satisfied. The latter ensure that the influence of the driving noise $\{e_k\}$ in the remote past on the present value of the process $\{a_k\}$ becomes vanishingly small. The stability conditions are expressed in terms of the autoregressive parameters only. Similarly, by recursive use of eqn (1), the ARMA(p,q) model can be transformed into a pure AR model of infinite order or ARMA$(\infty,0)$ provided the invertibility conditions are satisfied. The latter warrant that the influence of the past values of the process $\{a_k\}$ on the present one becomes smaller as we go further in the past. The inverse situation is physically meaningless and the invertibility conditions are usually imposed when modeling physical systems. The invertability conditions are expressed in terms of the moving average parameters only. If the stability conditions are satisfied, the output process $\{a_k\}$ has a finite variance and is a discrete stationary Gaussian process completely described by its mean function $(\mu_a = 0)$ and autocovariance or autocorrelation function in the time domain or its spectrum in the frequency domain.[5] The definitions of the time and frequency domain second-order statistical properties are given next:

Autocovariance function:

$$\gamma_n = \mathrm{E}\left[(a_k - \mu_a)(a_{k+n} - \mu_a)\right]$$

$$(2)$$

Autocorrelation function:

$$\mathscr{R}_n = \mathrm{E}\left[a_k a_{k+n}\right]$$

$$(3)$$

Variance:

$$\mathrm{Var}\left[a_k\right] = \sigma_a^2 = \gamma_0$$

$$(4)$$

Autocorrelation coefficient function:

$$\rho_n = \frac{\mathrm{E}\left[(a_k - \mu_a)(a_{k+n} - \mu_a)\right]}{\sigma_a^2} = \frac{\gamma_n}{\gamma_0}$$

$$(5)$$

One-sided spectrum:

$$p(f) = 2\sigma_e^2 \frac{\left|1 - \sum_{j=1}^{q} \theta_j e^{-i2\pi jf\Delta t}\right|^2}{\left|1 - \sum_{j=1}^{p} \phi_j e^{-i2\pi jf\Delta t}\right|^2} \Delta t, \quad (0 \leq f \leq f_{\mathrm{Nyq}})$$

$$(6)$$

In the above, E[.] represents the expectation operator and $i = \sqrt{-1}$. If $q - p < 0$, the autocorrelation function ρ_n consists of a mixture of damped exponentials and/or damped sine waves.[5] The spectrum defined in eqn (6) is the discrete counterpart of the power spectral density (PSD) function encountered in continuous-time random vibration theory. The mean square (or variance since

$\mu_a = 0$) of the ARMA process $\{a_k\}$ can be obtained from the one-sided spectrum:

$$\text{Var}\,[a_k] = \sigma_a^2 = \int_0^{f_{\text{Nyq}}} p(f)\,\mathrm{d}f \tag{7}$$

As indicated by eqn (7), the spectrum $p(f)$ has a physical interpretation as the distribution of the average energy (or mean square) of the process over the continuous range of frequency $[0, f_{\text{Nyq}}]$ where $f_{\text{Nyq}} = 1/(2\,\Delta t)$ stands for the Nyquist frequency or highest frequency which can be observed in the discrete time series $\{a_k\}$.

Physical interpretation

Under certain conditions, the ARMA processes can be interpreted as continuous-time response processes of dynamic systems uniformly sampled at time interval Δt. For example, it can be shown[12,19] that the response covariance function of a white-noise excited n-degree-of-freedom (n-dof) continuous-time linear dynamic system discretely coincides with the covariance function of an ARMA$(2n, 2n-1)$ discrete process. The parameters of the ARMA$(2n, 2n-1)$ model depend on the dynamic characteristics of the covariance-equivalent n-dof system, (i.e. natural periods and modal damping ratios) and the properties (temporal and spatial) of the white-noise excitation vector. This property represents an important advantage of using discrete ARMA models instead of continuous-time models for earthquake ground motion modeling, since discrete models are easier to identify from real digital seismic data and can still be interpreted in terms of continuous-time dynamic models. As an important illustration, the covariance equivalence existing between the low order ARMA$(2, 1)$ model and the well-known continuous Kanai-Tajimi model is presented in the next section.

ARMA$(2, 1)$ model

The ARMA$(2, 1)$ model is defined by the second order difference equation

$$a_k - \phi_1 a_{k-1} - \phi_2 a_{k-2} = e_k - \theta_1 e_{k-1} \tag{8}$$

and is completely characterized by four parameters: ϕ_1, ϕ_2, θ_1, and σ_e^2. The stability region for the autoregressive parameters ϕ_1 and ϕ_2 corresponds to the triangular region shown in Fig. 1[5] and the invertibility condition is satisfied if $|\theta_1| < 1$.

The underlying physical system corresponding to the ARMA$(2, 1)$ model is the linear viscously damped SDOF system represented in Fig. 2 where $m = $ mass of the SDOF oscillator, $k = $ linear spring stiffness, $c = $ dashpot damping coefficient, $X(t) = $ input displacement applied separately to the spring and the dashpot in proportions C_s and C_d, respectively, $Z(t) = $ absolute displacement of the SDOF system (measured with respect
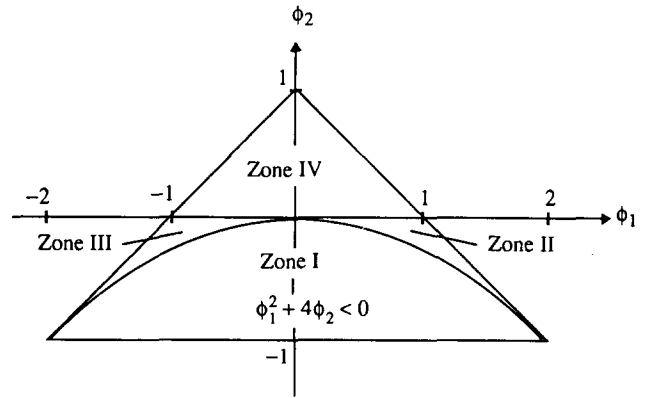

Fig. 1. Stability region.

to a fixed inertial coordinate frame). The equation of motion of the SDOF system can be written as:

$$-c\,[\dot{Z}(t) - C_d \dot{X}(t)] - k[Z(t) - C_s X(t)] = m\ddot{Z}(t) \tag{9}$$

By defining $\omega_g^2 = k/m$ and $2\xi_g \omega_g = c/m$, the equation of motion can be rewritten in the standard form:

$$\ddot{Z}(t) + 2\xi_g \omega_g \dot{Z}(t) + \omega_g^2 Z(t) = C_s \omega_g^2 X(t) + 2C_d \xi_g \omega_g \dot{X}(t) \tag{10}$$

It can be shown[8,10] that if the input acceleration process $\ddot{X}(t)$ is a continuous white-noise of constant power spectral density ϕ_0, then the continuous response process $a(t) = \ddot{Z}(t)$ is covariance equivalent (discretely coincident) with the discrete ARMA$(2, 1)$ process. For the continuous, as well as the discrete case, the autocovariance and autocorrelation functions (ACF) coincide since both input and output have a zero mean. The equivalence between the parameters of the discrete model $(\phi_1, \phi_2, \theta_1, \sigma_e^2)$ and the parameters of the continuous model $(\omega_g, \xi_g, C_s/C_d, \phi_0)$ is obtained by comparing and equating, at all discrete times $t_n = n\Delta t$, $n = 0, 1, 2, \ldots$, the autocorrelation functions of the discrete and continuous models. The latter are presented in Tables 1 and 2, respectively, where the autocorrelation function is expressed as the product of the variance (= mean-square value) and the autocorrelation coefficient function (ACCF) ρ_n or $\rho(\tau)$. In the case of the discrete ARMA$(2, 1)$ model, the ACF can
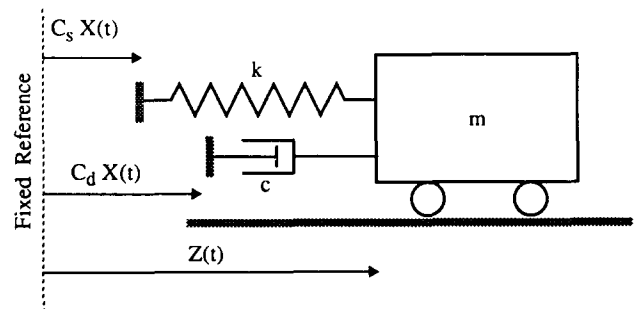

Fig. 2. Underlying physical system.

Table 1. Discrete autocorrelation coefficient function ρ_n

Zone	$\rho_n = \rho(n\Delta t),\ n = 0, 1, 2, \ldots$
I	$$\rho_n = e^{n/2 \ln(-\phi_2)}\frac{\cos(n\theta - \mu_d)}{\cos(-\mu_d)}$$ $$\mu_d = \arctan\left(\frac{2\rho_1 - \phi_1}{\sqrt{-(\phi_1^2 + 4\phi_2)}}\right), \quad \theta = \arccos\left(\frac{\phi_1}{2\sqrt{-\phi_2}}\right)$$ $$\rho_1 = \frac{\phi_1(1 + \theta_1^2 - \theta_1\phi_1) - \theta_1(1 - \phi_2^2)}{(1 - \phi_2)(1 + \theta_1^2 - \theta_1\phi_1) - \theta_1\phi_1(1 + \phi_2)}$$
II	$$\rho_n = e^{cn}\left[\cosh(\delta n) + \frac{e^{-c}\rho_1 - \cosh(\delta)}{\sinh(\delta)}\sinh(\delta n)\right]$$ $$c = \tfrac{1}{2}\ln(-\phi_2), \quad \delta = \tfrac{1}{2}\ln\frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}$$ ρ_1 as for Zone I
III	$$\rho_n = e^{cn}\left[(-1)^n \cosh(\delta n) + (-1)^{n+1}\frac{e^{-c}\rho_1 + \cosh(\delta)}{\sinh(\delta)}\sinh(\delta n)\right]$$ c, δ, and ρ_1 as for Zone II
IV	$$\rho_n = e^{cn}\left[\cosh(\delta n) + (-1)^{n+1}\frac{e^{-c}\rho_1 - \cosh(\delta)}{\sinh(\delta)}\sinh(\delta n)\right]$$ $$c = \tfrac{1}{2}\ln(\phi_2), \quad \delta = \tfrac{1}{2}\ln\left(-\frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}\right), \quad \text{and } \rho_1 \text{ as for Zone I}$$

Variance: $$E[a_k^2] = \sigma_a^2 = \frac{(1 - \phi_2)(1 + \theta_1^2) - 2\phi_1\theta_1}{(1 + \phi_2)[(1 - \phi_2)^2 - \phi_1^2]}\sigma_e^2 \quad \text{(Zones I, II, III, IV)}$$

Table 2. Continuous autocorrelation coefficient function $\rho(\tau)$

Underdamped case: $\xi_g < 1$

$$\rho(\tau) = e^{-\xi_g \omega_g |\tau|}\left[\frac{\cos(\omega_{gd}|\tau| - \mu_c)}{\cos(-\mu_c)}\right]$$

$$\omega_{gd} = \omega_g\sqrt{1 - \xi_g^2}, \quad \mu_c = \arctan\left[\frac{(C_s^2 - 4C_d^2\xi_g^2)}{(C_s^2 + 4C_d^2\xi_g^2)}\frac{\xi_g}{\sqrt{1 - \xi_g^2}}\right], \quad -\frac{\pi}{2} < \mu_c < \frac{\pi}{2}$$

Overdamped case: $\xi_g > 1$

$$\rho(\tau) = e^{-\xi_g\omega_g|\tau|}\left[\cosh(\omega_g\sqrt{\xi_g^2 - 1}|\tau|) + \frac{C_s^2 - 4C_d^2\xi_g^2}{C_s^2 + 4C_d^2\xi_g^2}\frac{\xi_g}{\sqrt{\xi_g^2 - 1}}\sinh(\omega_g\sqrt{\xi_g^2 - 1}|\tau|)\right]$$

Variance: $$E[a(t)^2] = \frac{\pi\phi_0}{2}\frac{\omega_g}{\xi_g}(C_s^2 + 4C_d^2\xi_g^2)$$

Table 3. Discrete/continuous parameter relationships for Zone I (Underdamped case: $\xi_g < 1$)

Continuous-to-discrete conversion: $\left(\omega_g, \xi_g, \dfrac{C_s}{C_d}\right) \rightarrow (\phi_1, \phi_2, \theta_1)$

$$\phi_1 = 2e^{-\xi_g \omega_g \Delta t} \cos\left(\omega_g \sqrt{1 - \xi_g^2}\,\Delta t\right)$$

$$\phi_2 = -e^{-2\xi_g \omega_g \Delta t}$$

$$\theta_1 = \text{solution of:} \quad \theta_1^2 + \frac{2\rho_1\phi_1 - \phi_1^2 + \phi_2^2 - 1}{\phi_1 - \rho_1(1 - \phi_2)}\theta_1 + 1 = 0,$$
$$|\theta_1| < 1$$

$$\rho_1 = \rho(\Delta t) = \frac{1}{2}\frac{\left(\dfrac{C_s}{C_d}\right)^2 - 4\xi_g^2}{\left(\dfrac{C_s}{C_d}\right)^2 + 4\xi_g^2}\frac{\xi_g}{\sqrt{1 - \xi_g^2}}\sqrt{-(\phi_1^2 + 4\phi_2)} + \frac{1}{2}\phi_1$$

$$0 < \omega_g\sqrt{1 - \xi_g^2} \le \frac{\pi}{\Delta t} = \text{Nyquist frequency}$$

$$\frac{(1 - \phi_2)(1 + \theta_1^2) - 2\phi_1\theta_1}{(1 + \phi_2)[(1 - \phi_2)^2 - \phi_1^2]}\sigma_e^2 = \frac{\pi\phi_0}{2}\frac{\omega_g}{\xi_g}(C_s^2 + 4C_d^2\xi_g^2)$$

Discrete-to-continuous conversion: $(\phi_1, \phi_2, \theta_1) \rightarrow \left(\omega_g, \xi_g, \dfrac{C_s}{C_d}\right)$

$$\omega_g = \frac{1}{2\Delta t}\sqrt{[\ln(-\phi_2)]^2 + 4\lambda_d^2}$$

$$\xi_g = \frac{-\ln(-\phi_2)}{\sqrt{[\ln(-\phi_2)]^2 + 4\lambda_d^2}}$$

$$\left(\frac{C_s}{C_d}\right)^2 = 4\xi_g^2\left(\frac{\xi_g + \sqrt{1 - \xi_g^2}\tan\mu_d}{\xi_g - \sqrt{1 - \xi_g^2}\tan\mu_d}\right)$$

$$\lambda_d = \arccos\left(\frac{\phi_1}{2\sqrt{-\phi_2}}\right), \quad 0 \le \lambda_d \le \pi$$

$$\mu_d = \arctan\left(\frac{2\rho_1 - \phi_1}{\sqrt{-(\phi_1^2 + 4\phi_2)}}\right), \quad -\frac{\pi}{2} \le \mu_d \le \frac{\pi}{2}$$

$$\rho_1 = \frac{\phi_1(1 + \theta_1^2 - \theta_1\phi_1) - \theta_1(1 - \phi_2^2)}{(1 - \phi_2)(1 + \theta_1^2 - \theta_1\phi_1) - \theta_1\phi_1(1 + \theta_2)}$$

Table 4. Discrete/continuous parameter relationships for Zone II (Overdamped case: $\xi_g > 1$)

Continuous-to-discrete conversion: $\left(\omega_g, \xi_g, \dfrac{C_s}{C_d}\right) \rightarrow (\phi_1, \phi_2, \theta_1)$

$$\phi_2 = -e^{-2\xi_g \omega_g \Delta t}$$

$$\phi_1 = -\frac{2}{\chi}\phi_2 + \frac{\chi}{2}, \quad \chi = 2e^{-\omega_g(\xi_g - \sqrt{\xi_g^2 - 1})\Delta t}$$

θ_1 as in Table 3

$$\rho_1 = \sqrt{-\phi_2}\left[\cosh(\delta) + \frac{\left(\dfrac{C_s}{C_d}\right)^2 - 4\xi_g^2}{\left(\dfrac{C_s}{C_d}\right)^2 + 4\xi_g^2}\frac{\xi_g}{\sqrt{\xi_g^2 - 1}}\sinh(\delta)\right]$$

$$\delta = \frac{1}{2}\ln\frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}$$

$$\frac{(1 - \phi_2)(1 + \theta_1^2) - 2\phi_1\theta_1}{(1 + \phi_2)[(1 - \phi_2)^2 - \phi_1^2]}\sigma_e^2 = \frac{\pi\phi_0}{2}\frac{\omega_g}{\xi_g}(C_g^2 + 4C_d^2\xi_g^2)$$

Discrete-to-continuous conversion: $(\phi_1, \phi_2, \theta_1) \rightarrow \left(\omega_g, \xi_g, \dfrac{C_s}{C_d}\right)$

$$\xi_g = \sqrt{\frac{[\ln(-\phi_2)]^2}{[\ln(-\phi_2)]^2 - 4\left[\text{arccosh}\left(\dfrac{\phi_1}{2\sqrt{-\phi_2}}\right)\right]^2}}$$

$$\omega_g = \frac{1}{2\Delta t}\sqrt{[\ln(-\phi_2)]^2 - 4\left[\text{arccosh}\left(\frac{\phi_1}{2\sqrt{-\phi_2}}\right)\right]^2}$$

$$\left(\frac{C_s}{C_d}\right)^2 = \frac{-4\xi_g^2\left[\left(\dfrac{\rho_1}{\sqrt{-\phi_2}} - \cosh(\delta)\right)\sqrt{\xi_g^2 - 1} + \xi_g\sinh(\delta)\right]}{\left(\dfrac{\rho_1}{\sqrt{-\phi_2}} - \cosh(\delta)\right)\sqrt{\xi_g^2 - 1} - \xi_g\sinh(\delta)}$$

ρ_1 as in Table 3 and δ as above

take four different forms depending on which subregion of the stability region the autoregressive parameters belong to. These four subregions are referred to as Zones I, II, III, IV, and are shown in Fig. 1. For the continuous model, both underdamped ($\xi_g < 1$) and overdamped ($\xi_g > 1$) cases have been considered. By comparing the mathematical expressions in Tables 1 and 2, it can be observed that coincidence of the continuous and discrete ACFs at all discrete times ($t_n = n\Delta t$, $n = 0, 1, 2, \ldots$) can exist only in Zones I and II. The terms $(-1)^n$ and $(-1)^{n+1}$ prevent that discrete coincidence in Zones III and IV. It is also noticed that Zones I and Zone II correspond to the underdamped and overdamped case, respectively. The one-to-one mapping between the continuous and

discrete parameters is summarized in Tables 3 and 4 for Zone I (underdamped case) and Zone II (over-damped case), respectively. This mapping depends on Δt, the sampling time interval. In the underdamped case, unicity of the discrete representation is warranted provided that Δt is sufficiently small, so that the damped natural frequency ($\omega_g\sqrt{1 - \xi_g^2}$) is smaller than the highest frequency, $1/2\Delta t$, known as the Nyquist frequency (f_{Nyq}). As indicated in Tables 3 and 4, the relationship between the variance of the input discrete white-noise and the power spectral density ϕ_0 of the continuous input white-noise is obtained by equating the variances of the discrete and continuous output processes.

The well-known Kanai-Tajimi (K-T) earthquake model[17,29] corresponds to the absolute acceleration response at ground level due to a white-noise acceleration input at bedrock level, the transfer between bedrock and ground levels being modeled as a linear viscously damped SDOF system. Therefore, the K-T stochastic earthquake model is covariance equivalent with the
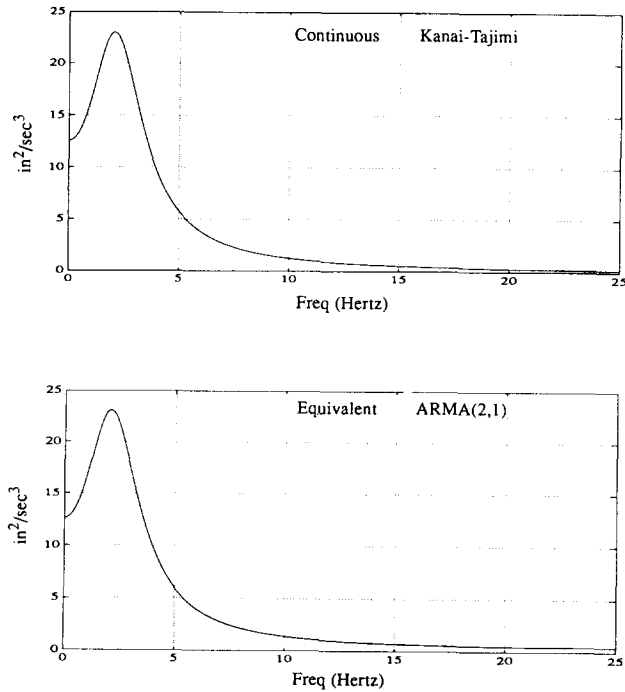
Continuous   Kanai-Tajimi

Freq (Hertz)

Equivalent   ARMA(2,1)

Freq (Hertz)

Fig. 3. Kanai-Tajimi and equivalent ARMA(2, 1) power spectra.

subset of the ARMA(2, 1) model corresponding to a unit value of the spring-dashpot input ratio ($C_s = C_d = 1$). For firm ground conditions, at moderate epicentral distance, Kanai and Tajimi have suggested the following parameter values: $\omega_g = 5\pi$ (or $T_g = 0.4$ sec) and $\xi_g = 0.6$. From the continuous-to-discrete conversion in Table 3, the corresponding ARMA(2, 1) parameters are:

$$(\omega_g = 5\pi, \ \xi_g = 0.60, \ C_s/C_d = 1.0, \ \phi_0 = 1.0)$$

$$(\phi_1 = 1.604, \ \phi_2 = -0.686, \ \theta_1 = 0.767, \ \sigma_e^2 = 39.08,$$

$$\Delta t = 0.02 \text{ sec})$$

The Kanai-Tajimi and ARMA(2, 1) power spectra corresponding to these numerical values are shown in Fig. 3. The K-T spectrum is defined for all frequencies ($0 \leq f \leq \infty$), whereas the ARMA(2, 1) spectrum is defined for frequencies in the Nyquist range ($0 \leq f \leq f_{Nyq}$). Therefore, if the value of the K-T power spectrum at the Nyquist frequency is small, the K-T and ARMA(2, 1) spectra practically coincide throughout the whole Nyquist range as in Fig. 3. Otherwise, the ARMA model reproduces the spectral density of the SDOF system only up to the point where the influence of the aliasing phenomenon becomes apparent.

Nonstationary case

Typical earthquake records exhibit important non-stationarities in amplitude (or intensity) and frequency content as emphasized by Liu.[20] Their taking into

consideration for seismic response prediction of structures is very important, although the nonstationarity in frequency content has often been disregarded in the past. In order to represent these two types of nonstationarity, a time-varying or dynamic version of the ARMA model of eqn (1) is chosen:

$$a_k - \phi_{1,k} a_{k-1} - \cdots - \phi_{p,k} a_{k-p}$$

$$= \sigma_k e_k - \theta_{1,k}(\sigma_{k-1} e_{k-1}) - \cdots - \theta_{q,k}(\sigma_{k-q} e_{k-q}) \quad (11)$$

where now $\{e_k\}$ is a unit-variance discrete Gaussian white-noise. The nonstationarity in amplitude is represented by the variance envelope $\{\sigma_{e,k}^2\}$ of the driving noise, whereas the nonstationarity in frequency content is modeled by the time-varying ARMA parameters $\phi_{i,k}$ and $\theta_{i,k}$. This uncoupling of the two types of non-stationarity is possible provided that the standard-deviation envelope $\{\sigma_k\}$ is slowly-varying in time compared to the periods of the earthquake process oscillations.

Motivated by the theoretical spectrum of a stationary ARMA process given in eqn (6), Kitagawa[18] defined the following instantaneous spectrum of a time-varying ARMA process by simply replacing the time-invariant ARMA coefficients with their time-dependent counterparts:

$$p(f, k) = 2\sigma_{e,k}^2 \frac{\left|1 - \sum_{j=1}^{q} \theta_{j,k} e^{-i2\pi jf\Delta t}\right|^2}{\left|1 - \sum_{j=1}^{p} \phi_{j,k} e^{-i2\pi jf\Delta t}\right|^2} \Delta t,$$

$$(0 \leq f \leq f_{Nyq}) \quad (12)$$

This instantaneous spectrum can be easily interpreted as a non-negative frequency-time decomposition of the expected earthquake energy. It is noteworthy that the above definition of the instantaneous power spectrum is not related to the classical instantaneous power spectrum defined by Page;[24] the latter is a partially negative frequency-time decomposition of the expected energy of a process.

Previous investigators have used time-varying linear time series models to represent and simulate strong earthquake ground motions. Determination of univariate and multivariate stationary ARMA models compatible with a given analytical power spectrum or spectral matrix was presented by Spanos[27] and Spanos & Mignolet.[28] Polhemus & Cakmak,[26] Chang et al.,[8] and Cakmak et al.[7] considered ARMA models modulated in amplitude using a variance envelope function. Jurkevics & Ulrych,[16] Gersch & Kitagawa,[13] Deodatis & Shinozuka,[11] and Papadimitriou[25] have used time-varying AR models, nonstationary in both amplitude and frequency content, estimated either directly from real earthquake accelerograms or from specified analytical evolutionary power spectra. In the present study, time-varying ARMA models are con-

sidered. They are more difficult to determine from earthquake records than AR models, but they have two important advantages compared to AR models: (i) they allow the modeling of sharp dips and sharp peaks of the ground motion power spectrum with fewer parameters since their transfer functions have both poles and zeros, unlike the all-pole AR models; and (ii) under certain conditions, they can be interpreted exactly as sampled output processes from white-noise excited multi-degree-of-freedom systems.

IDENTIFICATION OF TIME-VARYING ARMA MODELS

In this study, two methods were used to estimate the time variation of the parameters of a given ARMA model from an actual earthquake accelerogram. These two methods are described below.

Moving time-window technique and windowed-periodogram spectral estimation

The conventional method, called moving time-window technique, stems from the assumption that the non-stationary time series analysed is approximately stationary within a time-window of size $n_{wd}\Delta t$. Standard statistical methods from time series analysis can then be used to estimate the ARMA parameters characterizing this segment of accelerogram. The parameter estimates corresponding to a certain window position are attached to the center point of the window. This stationary parameter estimation is repeated for successive equi-distant window positions. In this study, the subroutine FTML from the IMSL library[15] has been used to perform the maximum likelihood estimation of the ARMA parameters and driving white-noise variance given a stationary time series. The time-window should be short enough to capture rapid changes in frequency content and long enough to provide for stable estimated parameters and the ability to capture low frequency components. From the experience gained in this research, it seems that a window size of 5·00 secs, corresponding to 251 ground acceleration values and a sampling time interval of 0·02 sec, realizes a good compromise.

The temporal variation of the frequency content of the earthquake record is estimated using the windowed-periodogram spectral estimation technique. The periodogram estimate of the power spectral density[6] of a stationary time series $\{a_k, k = 1, \dots, L\}$ is defined as:

$$P_k = P(\omega_k) = \frac{\Delta t}{2\pi L}|A(\omega_k)|^2 \qquad (13)$$

where $A(\omega_k)$ is the discrete Fourier transform of $\{a_k\}$:

$$A(\omega_k) = \sum_{n=0}^{L-1} a_n e^{-i2\pi nk/L}, \qquad k = 0, \dots, L-1 \qquad (14)$$

and $\omega_k = k\Delta\omega$, $\Delta\omega = 2\pi\Delta f$, $\Delta f = 1/L\Delta t$ = frequency resolution. From the discrete form of the Parseval's theorem which expresses that the total power in a signal is the same whether it is computed in the time domain or in the frequency domain, it follows that:

$$\sum_{k=0}^{L-1} P(\omega_k)\Delta\omega = 2\pi \sum_{k=0}^{L-1} P(f_k)\Delta f = \frac{1}{L}\sum_{k=0}^{L-1} a_k^2 \qquad (15)$$

For a zero-mean time series $\{a_k\}$, the right-hand-side of eqn (15) represents the sample variance. As in the case of a continuous-time stochastic process for which the variance equals the area under the process power spectral density and as in eqn (7), eqn (15) expresses that the area under the periodogram is equal to the variance of the time series and reveals how it is contributed at the various discrete frequencies $(f_k = k\Delta f, k = 0, \dots, L - 1)$. The discrete Fourier transform in eqn (14) is computed using a fast Fourier transform algorithm. To reduce the spectral leakage[6] between neighboring discrete frequencies, the Parzen taper[6] is applied to the segment of data $\{a_k, k = 1, \dots, L\}$ prior to computing their Fourier transform. In order to reduce the large variance of the periodogram estimate defined in eqn (13), spectral smoothing using the differential weighting [1/4, 1/2, 1/4] is applied. The windowed-periodogram spectral estimation technique consists in evaluating the periodogram over successive overlapping windows of the data $\{a_k\}$.

Iterative Kalman filtering

The basic idea of a time-adaptive algorithm is to optimally update the system parameters at each time step using the observation at this time step. In this manner, a continuous description of the time-varying properties of the input data is obtained. A technique of adaptive estimation based on Kalman filtering theory is used for the identification of the time-varying ARMA coefficients of the earthquake model represented by eqn (11). This technique represents an iterative version of a previous method developed by Nau *et al.*[21]

For application of the Kalman filter, the general dynamic ARMA(p, q) model represented by eqn (11) must be recast in the following discrete state-space form:

$$\boldsymbol{\xi}_{k+1} = \boldsymbol{\xi}_k + \boldsymbol{\delta}_k \qquad (16)$$

$$a_{k+1} = \mathbf{H}_{k+1}^T\boldsymbol{\xi}_{k+1} + e_{k+1} \qquad (17)$$

where

$$\boldsymbol{\xi}_k = [\phi_{1,k}, \dots, \phi_{p,k}, -\theta_{1,k}, \dots, -\theta_{q,k}]^T \qquad (18)$$

$$\boldsymbol{\delta}_k = [\delta_{1,k}, \dots, \delta_{p+q,k}]^T \qquad (19)$$

$$\mathbf{H}_k = [a_{k-1}, \dots, a_{k-p}, e_{k-1}, \dots, e_{k-q}]^T \qquad (20)$$

In the above equations, the superscript $(\ )^T$ indicates the transpose operation. The earthquake signal model is thus completely defined by eqn (16), called the transition (or state) equation and which governs the 'motion' in time of the state vector ξ_k, and by eqn (17), called the observation (or measurement) equation. The state vector ξ_k is partitioned in such a way that the first terms are the autoregressive parameters and the remaining terms the moving average parameters. The random vector δ_k represents the motion or dynamic noise, also called 'plant' noise in the control literature. The observation a_k represents the observed ground acceleration value at time $t_k = k\Delta t$ ($k = 0, 1, 2, \ldots$). The vector $\mathbf{H}_k$, called the 'history' vector, gives the ideal (noiseless) connection between the observation a_k and the state vector ξ_k at time k. Finally, $\{e_k\}$ represents the nonstationary Gaussian driving white-noise of the ARMA model. The following assumptions are made:

$$\mathrm{E}[\delta_k] = \mathbf{0}, \quad \mathrm{Cov}[\delta_k, \delta_k^T] = \mathbf{\Sigma}_\delta \tag{21}$$

$$\mathrm{E}[e_k] = 0, \quad \mathrm{Var}[e_k] = \sigma_{e,k}^2 \tag{22}$$

$$\mathrm{Cov}[\delta_k, e_k] = \mathbf{0}. \tag{23}$$

The random walk model for the parameter vector ξ_k adopted in eqn (16) may not appear physically reasonable. However, it is the simplest equation of motion that can be assumed in the absence of detailed prior information about parameter behavior. In this context, it is merely a device for ensuring that the Kalman filter will remain continually adaptive to any changes in the parameters that may occur. The covariance matrix $\mathbf{\Sigma}_\delta$ governs the average step size of the random walk assumed for the parameters and therefore it determines the long-run sensitivity of the filter to changes in the parameters. In this study, it has been assumed that $\mathbf{\Sigma}_\delta = \sigma_\delta^2 \mathbf{I}$, where σ_δ is a small number compared to average values of the parameters, (e.g. $\sigma_\delta \leq 0.01$). The problem of 'tuning' the filter to obtain parameter estimates which on one hand respond quickly to nonstationarities in the data, but on the other hand are not unstable, is difficult and has been tackled by trial-and-error in selecting the parameter σ_δ.

The discrete-time Kalman filter[2,10,21] provides a set of recursive equations to trace the evolution of the probability distribution of the state vector ξ_k in time. The expected value of ξ_k conditional on the observations up to time $k - 1$ and k are called the prior and posterior expected value of ξ_k; they are denoted by $\hat{\xi}_{k|k-1}$ and $\hat{\xi}_{k|k}$, respectively, and are used as estimates of the parameter vector ξ_k. Let $\mathbf{P}_{k|k-1}$ and $\mathbf{P}_{k|k}$ denote the corresponding estimation error covariance matrices defined by $\mathrm{E}[(\xi_k - \hat{\xi}_{k|\ldots})(\xi_k - \hat{\xi}_{k|\ldots})^T]$. At first, $\hat{\xi}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$ are computed using the predictor Kalman filter equations and then $\hat{\xi}_{k|k}$ and $\mathbf{P}_{k|k}$ are obtained using the updating Kalman filter equations. The presence of moving-average terms in the selected earthquake model introduces some nonlinearity in the filtering problem,

since products of unknown parameters and noise values appear in the observation, eqn (17). Therefore, some form of 'linearization' of the basic Kalman filtering algorithm is required. Note that, in practice, the true measurement noise $\{e_k\}$ is never observed directly. Hence, when implementing the Kalman filter, estimates of the true noise terms $(e_k, \ldots, e_{k-q+1})$ must be substituted in the 'history' vector $\mathbf{H}_{k+1}$. The residuals corresponding to the difference between prediction and observation, i.e. $\hat{e}_{k+1} = a_{k+1} - \mathbf{H}_{k+1}^T \hat{\xi}_{k+1|k+1}$, are used for this purpose. Furthermore, the variance envelope of the true driving noise $\{\sigma_{e,k}^2\}$ is needed as input to the Kalman filter. Because $\{\sigma_{e,k}^2\}$ is *a priori* unknown, the Kalman filter is applied iteratively using a succession of improved estimates of the true noise variance envelope. For the first iteration, a rough estimate of $\{\sigma_{e,k}^2\}$ is obtained by using a scaled version of the variance envelope estimate of the observed process (earthquake record). The variance envelope estimate of the residuals generated by the first application of the Kalman filter is then used as an estimate of the true noise variance envelope during the second application of the Kalman filter. The Kalman filter is applied iteratively until the difference between the assumed noise variance envelope and the variance envelope estimate of the residuals generated by the Kalman filter satisfies a given error tolerance. The non-parametric variance-envelope estimation technique developed by Nau *et al.*,[21] called the 'two-stage weighted moving average estimate', has been used in this study. The p initial digital ground acceleration values of the earthquake record analysed are used to build the initial 'history' vector $\mathbf{H}_p = [a_{p-1}, \ldots, a_0, 0, \ldots, 0]^T$ in which the estimates of the initial true noise terms are taken to be zero. Initial conditions for the state vector, $\hat{\xi}_{p-1|p-1}$, and the error covariance matrix, $\mathbf{P}_{k-1|k-1}$, are also required. An initial estimate for the state vector is obtained by performing a maximum likelihood estimation within a time window positioned at the beginning of the earthquake record. An initial error covariance matrix equal to the covariance matrix $\mathbf{\Sigma}_\delta$ of the random walk of the ARMA parameters is chosen. The first observed data used by the Kalman filter is a_p. From these initial conditions, the Kalman filter is run recursively until the last data element a_{N-1}.

MODEL VERIFICATION

If an ARMA model were to fit perfectly a real earthquake accelerogram, it would map the highly correlated time series corresponding to the actual accelerogram into a realization of a perfectly uncorrelated stationary Gaussian white-noise. Therefore, the first type of model diagnostic checking is applied to the final (converged) sequence of normalized residuals $\{\hat{e}_k/\hat{\sigma}_{e,k}\}$ given by the model estimation procedure.

These residuals are checked for 'whiteness' or uncorrelation and normality. However, a model is never perfect and in the present case, the imperfection can be attributed to the following reasons:

- The model estimation procedure generates inaccurate or incorrect results.
- The selected model order (p, q) is not adequate for the nature of the particular earthquake process considered.
- The assumption of a linear earthquake model is too restrictive given the various non-linearities associated with the actual physical earthquake process.

The next question to be addressed is whether the imperfect estimated earthquake model is sufficient for representation of the ground motion variability and the corresponding structural response variability associated with a given real earthquake process. To answer this question, the estimated earthquake model is used further for ground motion and structural response simulation.

The second type of model diagnostic checking is applied to individual or an ensemble of simulated accelerograms to verify the extent to which they resemble the actual 'target' accelerogram. For this purpose, the following checking operations are performed:

- Analyse ARMA-model-simulated accelerograms, utilizing the analysis and estimation techniques used on the 'target' accelerogram. The objective of this analysis is to verify the accuracy with which the estimation techniques can recover the 'true' parameters of the earthquake process when they are known.
- Compare ground velocity and displacement from simulated and 'target' motions. This test is used to insure that not only the simulated ground acceleration, but also the corresponding ground velocity and displacement time histories are realistic, since the latter two can control the earthquake response of certain structures.
- Compare the Fourier amplitude spectra of the simulated and 'target' ground motions (acceleration, velocity, and displacement). The Fourier amplitude spectra give an indication of the global frequency content over the whole duration of the earthquake.
- Compare the instantaneous spectrum of the fitted ARMA model to the windowed-periodogram estimate of the time-varying power spectral density characterizing the 'target' accelerogram.
- Compare the traditional ground motion parameter values of the 'target' earthquake to the statistics of these parameters computed from an ensemble of simulated earthquakes. The ground motion parameters considered include: peak ground acceleration (PGA), velocity (PGV) and displacement (PGD), root-mean-square acceleration (RMSA), velocity (RMSV) and displacement (RMSD), and Housner's spectral intensity (SI).

Residual checks

Checking of the normalized residuals $(w_k = \hat{e}_k / \hat{\sigma}_{e,k}$, $k = 0, \ldots, N-1)$ for uncorrelation ('whiteness') can be performed either in the time domain, based on the sample autocorrelation coefficient function, or in the frequency domain, based on the periodogram. The sample autocorrelation coefficient function (ACCF) at lag $(k\Delta t)$, $\hat{\rho}_k$, is defined as:

$$\hat{\rho}_k = \hat{R}_k / \hat{R}_0 \tag{24}$$

where

$$\hat{R}_k = \frac{1}{N} \sum_{i=0}^{N-k-1} (w_i - \bar{w})(w_{i+k} - \bar{w})$$

and $\bar{w} = (1/N)\Sigma_{k=0}^{N-1} w_k$. If $\{w_k\}$ is a purely random (uncorrelated) process, it can be shown[19] that:

$$\mathrm{E}\,[\hat{\rho}_k] \cong 0 \quad \text{and} \quad \mathrm{Var}\,[\hat{\rho}_k] \cong \frac{1}{N}, \quad \text{for all } k \neq 0 \tag{25}$$

Hence, on the basis of the asymptotic normal distribution of $\hat{\rho}_k$, an approximate 95% confidence interval for $\hat{\rho}_k$ is $(-2\sqrt{1/N}, 2\sqrt{1/N})$. Another time domain test for goodness of fit is provided by the so-called 'Portmanteau' statistic:[5]

$$Q = N \sum_{k=1}^{n} \hat{\rho}_k \tag{26}$$

where N is the number of data points, and $n \cong N/5$. Under the hypothesis that the residuals are completely uncorrelated, Q is approximately chi-square distributed with the number of degrees-of-freedom equal to n minus the number of estimated parameters, $(p + q)$. Therefore, an approximate 'Portmanteau' test of the hypothesis of model adequacy may be made by referring an observed value of Q to a table of the percentage points of χ^2. However, as a general rule, a value of Q not much larger than the number of DOFs is considered to indicate a good fit. In the frequency domain, visual inspection of the periodogram computed from the normalized residuals is used to check their whiteness. A non-uniform shape of the periodogram reveals the presence of correlation among the residuals.

The normality test of the normalized residuals is performed by visual inspection of their plot on normal probability paper. For this purpose, the mth value among N residuals (pre-sorted in increasing order) is plotted versus the standard normal value corresponding to the cumulative probability $m/(N + 1)$. A lack of linearity in the resulting graph of data points suggests that the underlying population is not Gaussian.

Simulation of artificial ground motions

To simulate artificial ground motions, the estimation or analysis procedure is simply reversed. The sequence of

operations performed is: (1) computer generation of a stationary, Gaussian, discrete white-noise; (2) time-modulation using $\hat{\sigma}_e(t)$; (3) ARMA filtering using $(\hat{\phi}_i(t),\ i = 1,\dots,p)$ and $(\hat{\theta}_i(t),\ i = 1,\dots,q)$; and (4) baseline correction using a high-pass Butterworth filter with cut-off frequency at about 0·1 Hz to eliminate part of the low frequency error introduced by the ARMA model. The discrete Gaussian white-noise is generated using the subroutine GGNML of the IMSL library.[15] Lack of cross-correlation between the members of an ensemble of artificial earthquake records simulated from the same estimated earthquake model is achieved by using a different seed number for each white-noise sequence generation.

Alternatively, artificial ground motions can be simulated based on specified time-varying physical properties of an underlying dynamic system such as the parameters $\omega_g(t)$, $\xi_g(t)$, and $(C_s/C_d)(t)$ for the ARMA(2, 1) model. The time-varying physical parameters are then converted into the corresponding time-varying ARMA parameters using the continuous-to-discrete parameter conversions. This approach can be generalized to multi-modal higher order ARMA $(2n, 2n - 1)$ models which are exactly covariance-equivalent with the sampled output of white-noise excited n-degree-of-freedom dynamic systems. This simulation procedure allows us to generate artificial ground motions com-patible with given evolutionary spectral properties described in terms of time-varying modal properties which are very meaningful to the dynamist.

APPLICATIONS

Applications of the stochastic model identification procedures described above are presented for two real earthquake accelerograms having very different characteristics. The first one corresponds to the north-south component of the moderate magnitude earthquake $(M_L = 6\cdot5)$, referred to as Event 39, recorded at the station E-02, classified as a rock site, of the SMART I accelerograph array in Taiwan.[1] The ground acceleration, velocity and displacement time histories of this seismic event are shown in Fig. 4. The ground acceleration record has been digitally recorded with a time sampling interval of 0·01 second. Hereafter, this record will be referred to as E02NS. The second target accelerogram corresponds to the north-south component of the 1940 Imperial Valley earthquake recorded at El Centro, California, on a deep firm alluvion deposit. The ground acceleration, velocity and displacement records are represented in Fig. 5. Only the first 35 seconds of the original record, digitized with a time sampling interval of 0·02 second, are considered in this study.

The parameters of the time-varying ARMA(2, 1) model identified from the E02NS record represented in
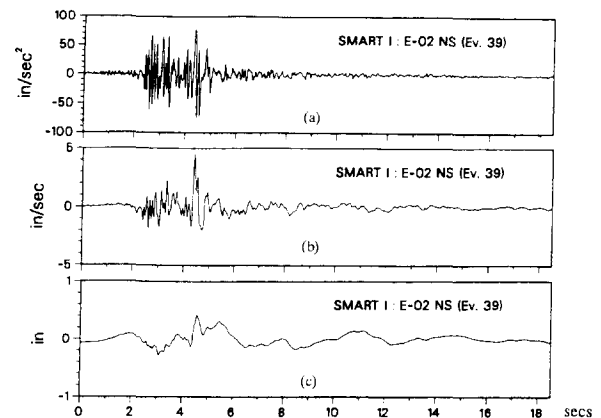

Fig. 4. (a) Ground acceleration, (b) velocity, and (c) displacement time histories of the E02NS record.

Fig. 4(a) using both the iterative Kalman filtering scheme and the moving window technique are displayed in Fig. 6. The results obtained using these two different methods agree overall. To evaluate the variance envelope by 'two-stage weighted moving average', a window size of $n = 30$ data points was used. For the Kalman filter, a trade-off value of $\sigma_\delta = 0\cdot008$ was selected by trial-and-error. A value of σ_δ which is too small produces a stable but too slow learning rate of the Kalman filter, which is unable to capture sharp changes in the structure of the earthquake signal. On the other hand, a value of σ_δ which is too large yields a fast learning rate, which results in very irregular and unrealistic time variations of model parameters. In applying the moving window technique, the following parameters were used: $n_{wd}\Delta t = 1\cdot0$ sec, time interval between successive window positions $= 0\cdot25$ sec. The convergence of the iterative Kalman filtering scheme is illustrated in Fig. 7 via the convergence of the standard deviation envelope estimate $\hat{\sigma}_{e,k}$ of the driving noise. In all cases considered in this study, a few iterations sufficed to obtain convergence. By using the discrete-to-
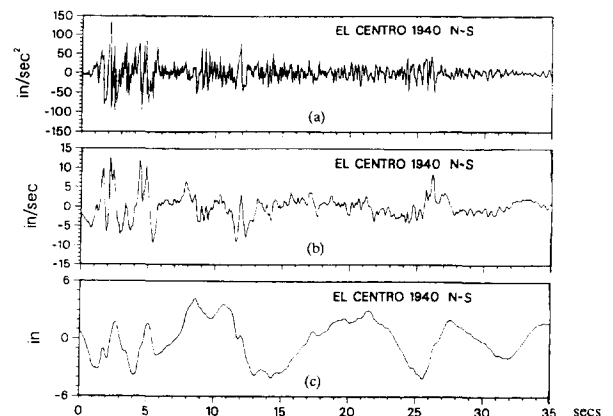

Fig. 5. (a) Ground acceleration, (b) velocity, and (c) displacement time histories of the El Centro 1940 record (N-S comp.).
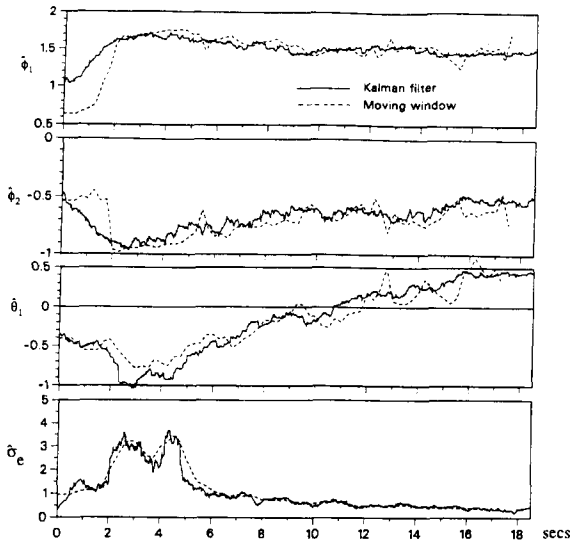
Fig. 6. ARMA(2, 1) model parameter estimates for the E02NS record.

Fig. 8. Time histories of physical parameters for the E02NS record.

continuous conversion relationships summarized in Tables 3 and 4, the time histories of the physical parameters corresponding to the ARMA(2, 1) model parameters of Fig. 6 have been computed and are depicted in Fig. 8. From Fig. 8(a), it can be seen that the predominant frequency F_g of the ground acceleration drops from 14 Hz to 7 Hz during the first 4 seconds of ground shaking, oscillates with an average value between 6 and 8 Hz in the middle portion of the ground motion, and finally drops to an average of 4 Hz at the end of the earthquake. In Fig. 8(b), the time history of the damping ratio parameter ξ_g indicates that the frequency bandwidth of the ground acceleration decreases during the first 3 seconds and then increases almost monotonically until the end of the record. The underlying SDOF model becomes overdamped ($\xi_g > 1$) only in the last few seconds of the record. The evolution in time of the input ratio parameter 'sign $((C_s/C_d)^2)\sqrt{|C_s/C_d|^2}$|' is displayed in Fig. 8(c). It is observed that the input of the
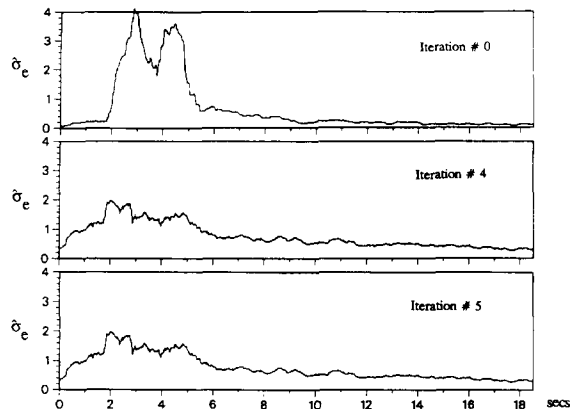
underlying SDOF model is imaginary $((C_s/C_d)^2 < 0)$ during the first 8 seconds and then becomes real-valued with the ratio C_s/C_d taking a large value ($\gg 1$), which means physically that the dashpot support of the SDOF model shown in Fig. 2 is almost fixed.

The time history of the normalized (amplitude-demodulated) residuals produced by the fitting of the ARMA(2, 1) model is shown in Fig. 9(a). The residuals sample ACCF $\hat{\rho}_k$ calculated up to a lag-time equal to one-fifth of the duration of the E02NS record is plotted in Fig. 9(b). The periodogram of the residuals obtained using a Parzen taper and without smoothing is displayed in Fig. 9(c). Figure 10 is similar to Fig. 9, but corresponds to an ARMA(4, 2) model fitted to the E02NS record. The two parallel dashed lines present in Figs 9 and 10 define the approximate 95% confidence interval for the sample ACCF $\hat{\rho}_k$ under the hypothesis that the residual process is purely random as indicated in eqn (25). It is observed from Fig. 9(b) that the residuals generated by the ARMA(2, 1) model fitted to the E02NS record possess some correlation structure at
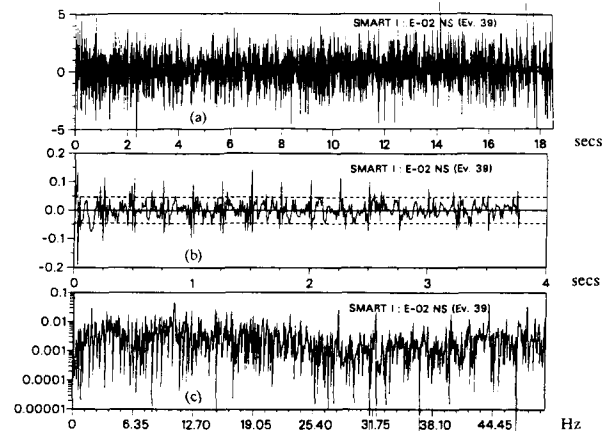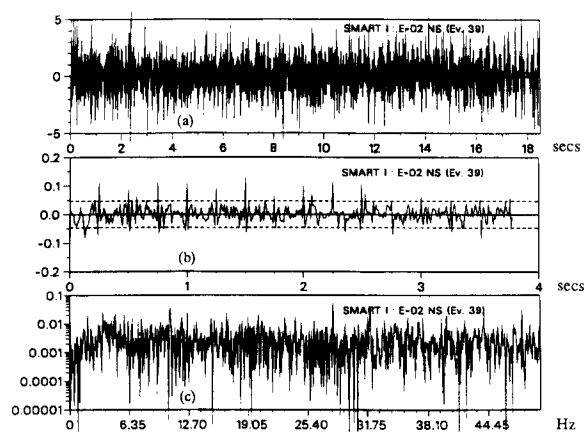

Fig. 7. Convergence of the iterative Kalman filtering scheme.

Fig. 9. Normalized residuals from the ARMA(2, 1) model fitted to the E02NS record: (a) time history, (b) sample autocorrelation function, and (c) periodogram.

Fig. 10. Normalized residuals from the ARMA(4,2) model fitted to the E02NS record: (a) time history, (b) sample autocorrelation function, and (c) periodogram.

small time-lags ($\tau < 0.20$ sec). Figure 9(c) reveals a non-uniformity of the power spectral density of the underlying residual process characterized by a partial spectral gap between 25 Hz and 40 Hz. By comparing Figs 9 and 10, it appears that the ARMA(4,2) model is better than the ARMA(2,1) model since its residuals are more uncorrelated. The results of the Portmanteau and '95% confidence interval' statistical tests are summarized in Table 5 for the residuals of the ARMA(2,1) and ARMA(4,2) models as well as for three computer-generated discrete Gaussian white-noises of the same duration (length). These results confirm the superiority of the ARMA(4,2) model over the ARMA(2,1) model since the former model has a smaller Portmanteau statistic value and a smaller percentage of $\hat{\rho}_k$ outside the 95% confidence interval. To quantify the uncorrelation

or whitening effect of the ARMA filters, the statistical test values are also reported in Table 4 for the amplitude-demodulated E02NS record obtained by scaling the original E02NS record by the inverse of its estimated standard-deviation envelope. The same statistical tests have been performed on the residuals generated from fitting an ARMA(2,1) model to the El Centro ground acceleration record (see Fig. 5(a)), and are reported in Table 6. From these results, it is observed that the representation of the El Centro record by an ARMA(2,1) model is superior to the representation of the E02NS record by either an ARMA(2,1) or an ARMA(4,2) model. The normality tests performed on the normalized residuals generated from fitting an ARMA(2,1) model to the E02NS record and to the El Centro record are also represented in Figs 11(a) and (b). In the case of the E02NS record, the linearity of the cumulative distribution of the residuals indicates that the underlying population of residuals is well represented by the Gaussian distribution. The same was observed for the ARMA(4,2) model. However, in the case of the El Centro record, the underlying population of residuals is not Gaussian as shown in Fig. 11(b).

The normalized ARMA time-varying power spectral densities derived from the Kalman filter results are compared to the normalized moving periodogram estimate of the time-varying frequency content of the E02NS record. This comparison is shown in Fig. 12. It is observed that when a single predominant peak is present in the periodogram estimate, it is captured accurately by both the ARMA(2,1) and ARMA(4,2) models, see Figs 12(a), (b), and (c). On the other hand, when several spectral peaks of the same magnitude are present in the periodogram estimate, both the ARMA(2,1) and

Table 5. Statistical tests for the ARMA models fitted to the E02NS record

Model	Q	# DOFs	Percentage of $\hat{\rho}_k$ outside 95% conf. interval (%)
ARMA(2,1)	883	375	16·1
ARMA(4,2)	567	372	9·3
Discrete white-noise #1	293	378	3·2
Discrete white-noise #2	334	378	1·9
Discrete white-noise #3	358	378	4·5
Ampl.-demod. E02NS	4774	370	

Table 6. Statistical tests for the ARMA model fitted to the El Centro 1940 record

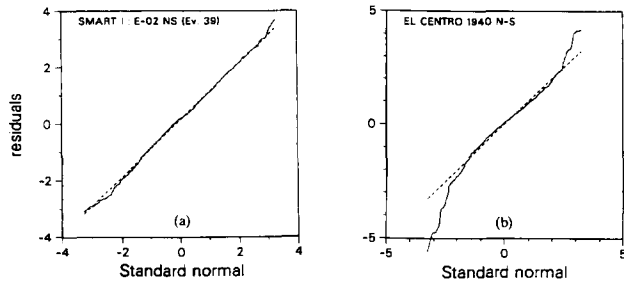| Model	Q	# DOFs	Percentage of $\hat{\rho}_k$ outside 95% conf. interval (%)
ARMA(2,1)	447	347	6·8
Discrete white-noise #1	273	350	2·6
Discrete white-noise #2	311	350	2·9
Discrete white-noise #3	322	350	4·6
Ampl.-demod. El Centro	4961	350	

Fig. 11. Normality test of the normalized residuals corresponding to the ARMA(2, 1) models fitted to (a) the E02NS record and (b) the El Centro 1940 record.
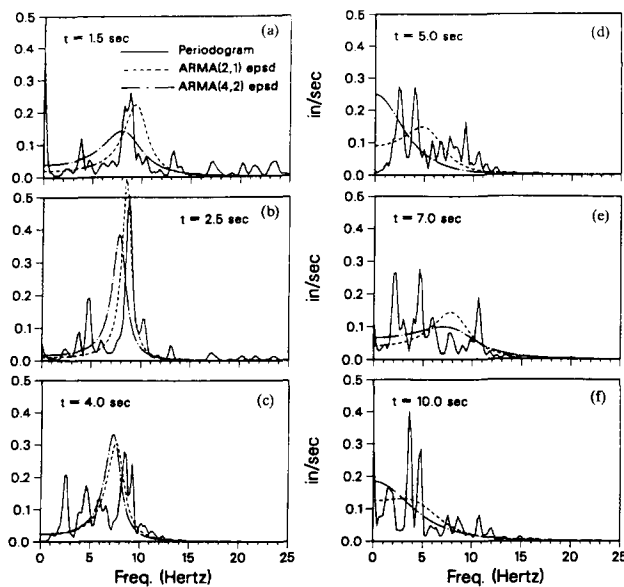

Fig. 12. Comparison between normalized time-varying ARMA power spectral estimate and normalized moving periodogram for the E02NS earthquake process.

ARMA(4, 2) time-varying power spectral densities exhibit a single peak whose position and shape can be interpreted as some weighted average of the positions and shapes of the spectral peaks indicated by the periodogram (see Figs 12(d), (e), and (f)). Even though

the ARMA(4, 2) model is bi-modal (can develop two spectral peaks), its Kalman filter estimate, in this particular example, was unimodal at all times. The graphic representation in frequency-time of the normalized ARMA(2, 1) power spectral density estimates for both the E02NS and the El Centro records are displayed in Fig. 13. These 3-D plots indicate the temporal variation of the frequency content of the ground shaking, commonly called frequency nonstationarity.

Artificial ground acceleration, velocity, and displacement corresponding to a single realization of the ARMA(2, 1) model fitted to the E02NS and El Centro records are shown in Figs 14 and 15, respectively. It is observed that the nonstationary amplitude and frequency characteristics of the target records are well captured by the ARMA(2, 1) models. However, from the artificial ground velocity and displacement it is observed that the very low frequency content of the artificial ground motions remains too high in spite of the high-pass Butterworth filtering applied with a cut-off frequency at about 0·10 Hz. This can be further corrected by increasing slightly the value of the cut-off frequency. The ARMA-simulated accelerogram represented in Fig. 14(a) has been analysed, utilizing the same iterative Kalman filtering scheme which was used to estimate the ARMA model from the target record. The results are presented in Fig. 16 and show that the 'true' model parameters shown in Fig. 6 can be recovered relatively well from an artificial realization.

Statistics of ground motion parameters

Using a sample of 100 artificial earthquake ground motions, the second order statistics (mean and coefficient-of-variation, c.o.v. = standard deviation/mean) of traditional ground motion parameters have been computed. These ground motion parameters include: peak ground acceleration (PGA), velocity (PGV), and displacement (PGD), root-mean-square acceleration (RMSA), velocity (RMSV), and displacement (RMSD),
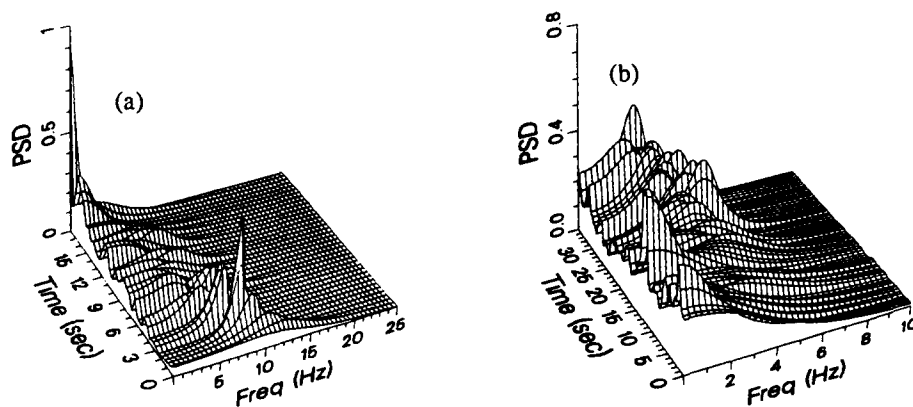

Fig. 13. Normalized ARMA(2, 1) time-varying PSD estimate for (a) the E02NS record, and (b) the El Centro record.
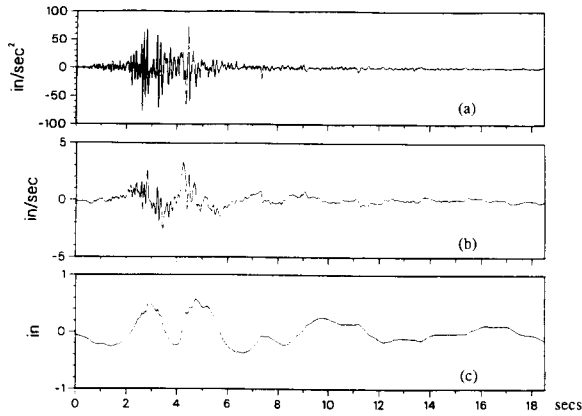
Fig. 14. ARMA(2, 1) artificial earthquake ground (a) acceleration, (b) velocity, and (c) displacement corresponding to the E02NS record.
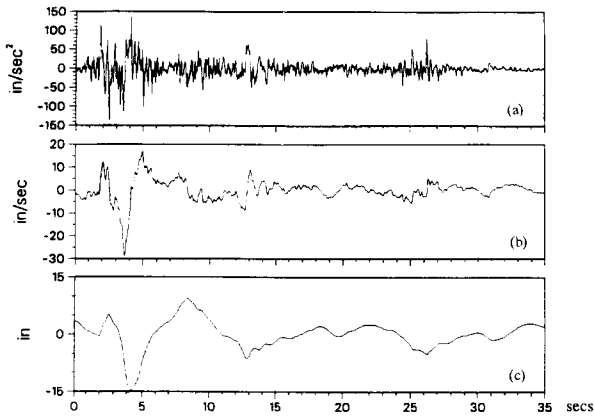

Fig. 15. ARMA(2, 1) artificial earthquake ground (a) acceleration, (b) velocity, and (c) displacement corresponding to the El Centro 1940 record.
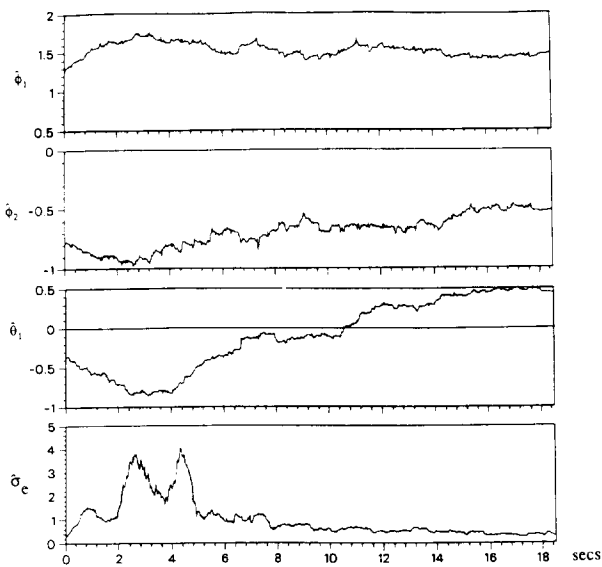

Fig. 16. ARMA(2, 1) model parameter estimates for an ARMA(2, 1) artificial realization of the E02NS earthquake process.

and Housner spectral intensity[†] (SI). These statistical results are summarized in Table 7 for the ARMA(2, 1) models fitted to the E02NS and El Centro records. In the same table, they are compared with the deterministic ground motion parameter values corresponding to the target records, and it can be checked that the statistical interval 'mean ± standard deviation' does not contain the parameter value of the target record in the case of PGV, PGD, RMSV, and RMSD due to the low frequency error of the ARMA model already mentioned. The statistics of PGA, RMSA, and SI agree with the corresponding deterministic ground motion parameter values of the target records. From Table 7, it is observed that for a given earthquake process, the RMSA parameter is the least variable among the set of ground motion parameters considered. In particular, RMSA is found much less variable than PGA, which has been used extensively in the past and is still used as an indicator of damage potential for earthquakes. The choice of a ground motion parameter characterized by a large variability and obtained from a single earthquake realization is not reliable. Statistical analyses of the simulated ground motion parameters showed that their statistical distribution is well approximated by lognormal distributions.

CONCLUSION

A stochastic system identification procedure based on iterative Kalman filtering is presented to estimate nonstationary ARMA models from actual earthquake accelerograms. Special attention is given to the ARMA(2, 1) model, a subset of which is covariance equivalent with the well-known continuous-time Kanai-Tajimi stochastic earthquake model. The complete set of parameter relationships between the discrete ARMA(2, 1) and the continuous K-T models is presented. Therefore, the discrete-to-continuous conversions allow the use of the proposed identification scheme to estimate the time-varying physical parameters of a K-T model non-stationary in both amplitude and frequency content. The irregular time histories of the resulting K-T model parameters can then be smoothened by least square fit of analytical functions before the nonstationary estimated K-T model can be used in analytical random vibration studies. Alternatively, the time histories of the physical parameters of the K-T model can be specified from engineering judgement and used as such in analytical random vibration studies or they can be converted into the time-varying parameters

[†] $SI = \int_{0 \cdot 10}^{2 \cdot 50} PSV(\xi, T) \, dT$, where $PSV(\xi, T)$ represents the pseudo-spectral velocity of the earthquake record for an SDOF oscillator of damping ratio $\xi = 0 \cdot 05$ and undamped natural period T.

Table 7. Statistics of ground motion parameters from ARMA(2, 1) earthquake models

Parameter	E02NS			El Centro 1940		
	Target	Mean	c.o.v.	Target	Mean	c.o.v.
PGA (in/sec²)	78·26	70·76	0·19	134·52	136·32	0·16
PGV (in/sec)	4·50	4·35	0·25	13·17	21·45	0·25
PGD (in)	0·42	2·34	0·48	4·28	14·02	0·36
RMSA (in/sec²)	11·90	9·61	0·12	21·80	20·76	0·07
RMSV (in/sec)	0.49	0·95	0·32	3·01	5·03	0·22
RMSD (in)	0·10	0·85	0·46	2·09	4·63	0·34
SI (in)	8·27	10·95	0·24	53·43	58·68	0·20

of the equivalent ARMA(2, 1) model using the continuous-to-discrete parameter conversions if Monte Carlo simulation studies are preferred.

Results of the application of the proposed identification procedure for two actual earthquake records are presented. The model validation tests performed are based on examination of the residuals, simulated ground motions (acceleration, velocity, and displacement), temporal variation of the ground motion frequency content, and statistics of traditional ground motion parameters.

It appears that time-varying ARMA models can be estimated very efficiently using iterative Kalman filtering and that they are able to capture well the nonstationary features of actual earthquake ground motions, especially the nonstationarity in frequency content. The latter, often neglected in past modeling of earthquake ground motions, can have very important effects on the seismic response of structures exhibiting inelastic degrading behavior.

These realistic ARMA ground motion models have been used to simulate the response of linear and nonlinear structures in order to test the validation of the earthquake model at the structural response level and to investigate how the structural response is influenced by ground motion and structural parameters when the inherent uncertainty contained in the earthquake ground motion is accounted for.[10]

ACKNOWLEDGEMENTS

REFERENCES

1. Abrahamson, N.A. *et al.* The SMART I accelerograph array (1980–1987): a review, *Earthquake Spectra*, 1987, **3**(2), 263–287.
2. Anderson, B.D.O. & Moore, J.B. *Optimal Filtering*, Prentice-Hall, New Jersey, 1979.
3. Anderson, J.C. & Bertero, V.V. Uncertainties in establishing design earthquakes, *Journal of the Structural Engineering Division*, ASCE, 1987, **113**(8), 1709–1725.
4. Araya, R. & Der Kiureghian, A. Seismic hazard analysis: improved models, uncertainties and sensitivities, Report No. UCB/EERC-90/11, Earthquake Engineering Research Center, University of California, Berkeley, CA, 1990.
5. Box, G.E.P. & Jenkins, G.M. *Time Series Analysis: Forecasting and Control*, revised edition, Holden-Day, San Francisco, 1981.
6. Brillinger, D. *Time Series — Data Analysis and Theory*, expanded edition, Holden-Day, San Francisco, 1981.
7. Cakmak, A.S., Sherif, R.I. & Ellis, G.W. Modeling earthquake ground motions in California using particular time-series method, *International Journal of Soil Dynamics & Earthquake Engineering*, 1985, **4**(3), 124–131.
8. Chang, M.K., Kwiatkowski, J.W. & Nau, R.F. ARMA models for earthquake ground motions, *International Journal of Earthquake Engineering & Structural Dynamics*, 1982, **10**, 651–662.
9. Clough, R.W. & Penzien, J. *Dynamics of Structures*, McGraw-Hill Book Co., Inc., New York, 1975.
10. Conte, J.P., Pister, K.S. & Mahin, S.A. Influence of the earthquake ground motion process and structural properties on response characteristics of simple structures, Report No. UCB/EERC-90/08, Earthquake Engineering Research Center, University of California, Berkeley, CA, 1990.
11. Deodatis, D. & Shinozuka, M. Autoregressive model for nonstationary stochastic processes, *Journal of the Engineering Mechanics Division*, ASCE, 1988, **114**(11), 1995–2012.
12. Gersch, W. & Liu, R. S-Z. Time-series methods for the synthesis of random vibration systems, *ASME Trans. Appl. Mech.*, 1976, 159–165.
13. Gersch, W. & Kitagawa, G. A time varying AR coefficient model for modelling and simulating earthquake ground motion, *International Journal of Earthquake Engineering & Structural Dynamics*, 1985, **13**, 243–254.
14. Housner, G.W. Statistics of pulses on strong-motion accelerograms, N.S.F. Seminar-Workshop on Strong Ground Motion, San Diego, California, 1978.
15. *IMSL The International Mathematical and Statistical Library, User's Manual*, version 1.1, January 1989.
16. Jurkevics, A. & Ulrych, T.J. Representing and simulating strong ground motion, *Bulletin of the Seismological Society of America*, 1978, **68**(3), 781–801.
17. Kanai, K. Semi-empirical formula for seismic characterization of the ground, *Bulletin of Earthquake Research Institute*, University of Tokyo, Tokyo, Japan, 1967, **35**.
18. Kitagawa, G. Changing spectrum estimation, *Journal of Sound & Vibration*, 1983, **89**, 433–445.
19. Kozin, F. & Natke, H.G. System identification techniques, *Structural Safety*, 1986, **3**, 269–316.

20. Liu, S.C. Evolutionary power spectral density of strong motion earthquakes, *Bulletin of the Seismological Society of America*, 1970, **60**(3), 891–900.
21. Nau, R.F., Oliver, R.M. & Pister, K.S. Simulating and analysing artificial nonstationary earthquake ground motions, *Bulletin of the Seismological Society of America*, 1982, **72**, 615–636.
22. Newmark, N.M. & Hall, W.J. Procedures and criteria for earthquake resistant design, *Building Practices for Disaster Mitigation*, National Bureau of Standards, Washington, D.C., Building Sciences Series 46, 1973, **1**, 209–236.
23. Newmark, N.M. & Riddell, R. Inelastic spectra for seismic design, *Proc. 7th World Conf. on Earthquake Engineering*, Vol. 4, Istanbul, Turkey, 1980, pp. 129–136.
24. Page, C.G. Instantaneous power spectra, *Journal of Applied Physics*, 1952, **23**(1), 103–106.
25. Papadimitriou, K. Stochastic characterization of strong ground motion and applications to structural response, Report No. EERL 90-03, California Institute of Technology, Pasadena, CA, 1990.
26. Polhemus, N.W. & Cakmak, A.S., Simulation of earthquake ground motion using ARMA models, *International Journal of Earthquake Engineering & Structural Dynamics*, 1981, **9**, 343–354.
27. Spanos, P.D. ARMA algorithms for ocean wave modeling, *Journal of Energy Resources Technology*, ASME, 1983, **105**, 300–309.
28. Spanos, P.D. & Mignolet, M.P. Simulation of stationary random processes: two-stage MA to ARMA approach, *Journal of the Engineering Mechanics Division*, ASCE, 1990, **116**(3), 620–641.
29. Tajimi, H. A standard method of determining the maximum response of a building structure during an earthquake, *Proc. 2nd World Conf. on Earthquake Engineering*, Vol. 2, Tokyo, Japan, 1960, pp. 781–798.