ANALYSIS AND MODELING OF EARTHQUAKE DATA

by

F. Kozin[I] and R. Gran[II]

INTRODUCTION

The general problem of statistical characterization of earthquake records remains, even to this day, an open field. It is almost incredible to realize that only recently have earthquake records been filtered to remove the dynamics of the accelerometers as well as other devices used in obtaining them. It is equally interesting to note at the micro-zonation meeting held in Seattle, November 1971, only one paper was devoted to a true statistical analysis of real earthquakes. The characterizations of earthquake records of the past have been based mainly upon various applications of Biot's ideas of a velocity spectrum, with newer modeling ideas coming from variety of investigators such as, Bogdanoff, Shinozuka, Caughey, Amen, Ang, and Liu as well as many others [2]. But what characterizes all of these works is the fact that the actual data record is not used to construct the model of the earthquake as a stochastic process.

There is no question that the modeling and characterization of earthquake records is a difficult problem. It is after all a non-stationary random phenomenon which, relatively speaking, does not occur too often in time at the place where recording devices are present.

Hence, we have had in the past a small number of records from which to determine statistical properties. Furthermore, investigators in the earthquake engineering field generally were not in contact with the vast advances in statistical estimation and modeling techniques that were motivated by the pressing data processing needs of the space program-especially the requirements for optimal control and trajectory analysis. Hence, modern approaches have not as yet occurred in the earthquake modeling field. It is now the time to look at earthquake records via these new powerful techniques, that have been so extremely useful in modern control and communications applications. The results to be gained from such a study should result in a more realistic model of earthquakes as random processes recursively estimated from the earthquake record, and being able to account immediately, in the estimation technique, for the fact that the observed record is noisey and has been filtered. The model to be estimated is in a form directly applicable to being used as inputs for similations purposes. Furthermore, it can easily be studied for various statistical properties such as axis crossings, frequency component properties, intensities, etc., etc. It can also be used in judging the merits of various other statistical models that have been proposed.

In the long run, however, there is an even more fascinating possibility to contemplate. The estimated model parameters could perhaps be associated with "neighboring" geological characteristics that could aid in the characterization process for micro-zonation.

[I] Professor, Electrical Engineering/Electrophysics Department, Polytechnic Institute of Brooklyn, Brooklyn, New York.

[II] Research Scientist, Research Department, Grumman Aerospace, Bethpage, New York.

PROBLEM FORMULATION

The techniques available for handling data in the present day can broadly be classified as correlation, spectral, maximum likelihood as well as Bayesian. The approach we use here is based upon the modern theory of optimal non-linear filtering, or recursive estimation, a Bayesian approach. It is unique in that it can be applied to non-stationary phenomena, and it must always converge to the minimum variance estimate. That is, no other estimation procedure can result in a smaller error, for the model that we are fitting to the observed data. Indeed, there are very few approaches that can handle the non-stationarity of the data. It is this point that has been the major stumbling block in the past.

We assume a model of the form

$$x^{(n)} + a_{n-1}(t)\, x^{(n-1)} + \cdots + a(t)\dot{x} + a_o(t)x = bW, \tag{1}$$

where W is the gaussian white noise. Thus, we assume from the outset that we are fitting the observed data to a gaussian process model .

The coefficients will be assumed to be of the form

$$a_i(t) = a_i^o + a_i^1 t + \cdots + a_i^{m_i} \frac{t^{m_i}}{m_i!} , \tag{2}$$

where the number of terms taken in the polynomial for each coefficient, may be dependent upon the coefficient. This must be assumed beforehand. Clearly, other coefficients can possibly be assumed, but this strikes us to be the easiest and obvious choice for this first look at the application.

The observed data is assumed to be of the form,

$$y = x(t) + n(t),$$

where $n(t)$ is the noise in the observations. This leaves us with unknown constants (a_i^k) to be estimated. The unknown constants are introduced as components of a new state vector z defined by

$$\begin{cases} z_j = \dfrac{d^{j-1}}{dt^{(j-1)}}\, x , \quad j = 1, \ldots, n . \\ \dot{z}_{n+i(k+1)} = Z_{n+ik} = \dfrac{d^k}{dt^k}\left(a_i^1 + a_i^2 t + \cdots + a_i^m \dfrac{t^{m-1}}{(m-1)!}\right) \end{cases} \tag{3}$$

$$i = 1, \ldots, n , \quad k = 1, \ldots, m .$$

The problem of identification is now formulated as optimal state estimation for the system

$$\begin{cases} z = f(z, t) + \sigma_1(z, t)w \\ y = h(z, t) + \sigma_2(y, t)n , \end{cases} \tag{4}$$

$$f(z, t) = A(z)z, \quad h(z, t) = Mz .$$

But, because the a_i^k's are coefficients of the derivatives of x , then the vector function $f(z, t)$ in (4) is non-linear and non-linear filtering ideas must be applied. In general the solution to a non-linear optimal filtering problem requires solving a complex stochastic partial differential equation. As the general solution to this partial differential equation is not known at this time, finite term approximations must be used. The so-called "extended" Kalman-filter approximation used successfully in aerospace applications does not work for our situation since the initial perturbations are large.

A new approximation to the optimal non-linear filter is applied In this case the conditional density, given the observed data is assumed to be symmetric about the origin. It follows that all odd moments are zero. This yields the optimal filter equation [1].

$$d\hat{z} = \hat{f}(z)dt + \hat{P}M'[\sigma_2\sigma_2']^{-1}[dy - M\hat{z}dt] \qquad (5)$$

where f, M, σ_2, y are defined in (4) and P is the conditional covariance matrix and "$\wedge$" indicates conditional means.

An application of this approach to El Centro South is shown in Fig. 1.

DISCUSSION

In the estimation procedure it is important that the so-called innovations process $dy-M\hat{z}dt$ is white noise. A number of tests such as the run test and the trend test were applied during the estimation procedure. In Fig. I the simulated time series generated by the derived model, using the estimated constants given in Fig. I, is plotted with the real earthquake record. As can be seen the general characteristics of the simulated earthquake match the actual earthquake very well. This is but one example of a number that has been calculated.

The time varying damping as well as the time varying undamped natural frequency terms are shown in Fig. II. For this group of estimated parameters the damping ratio decreases then increases linearly. Similar properties are illustrated in the undamped natural frequency as well.

This work is merely a preliminary study of the application of modern filtering techniques to the problem of analyzing and characterizing earthquake records. It is our sincere feeling that this approach will become very useful in future earthquake engineering studies.

[1] R. Gran; System Identification using Approximate Non-Linear Filters, Proc. Third Symp. on Nonlinear Estimation, Sept. 1972 IEEE No. 72C.

[2] R. Levy, F. Kozin, R.B. Moorman; Random Processes for Earthquake Simulation, Jour. Eng. Mech. Div.; ASCE 97, EMZ April 1971, pp. 495-517.

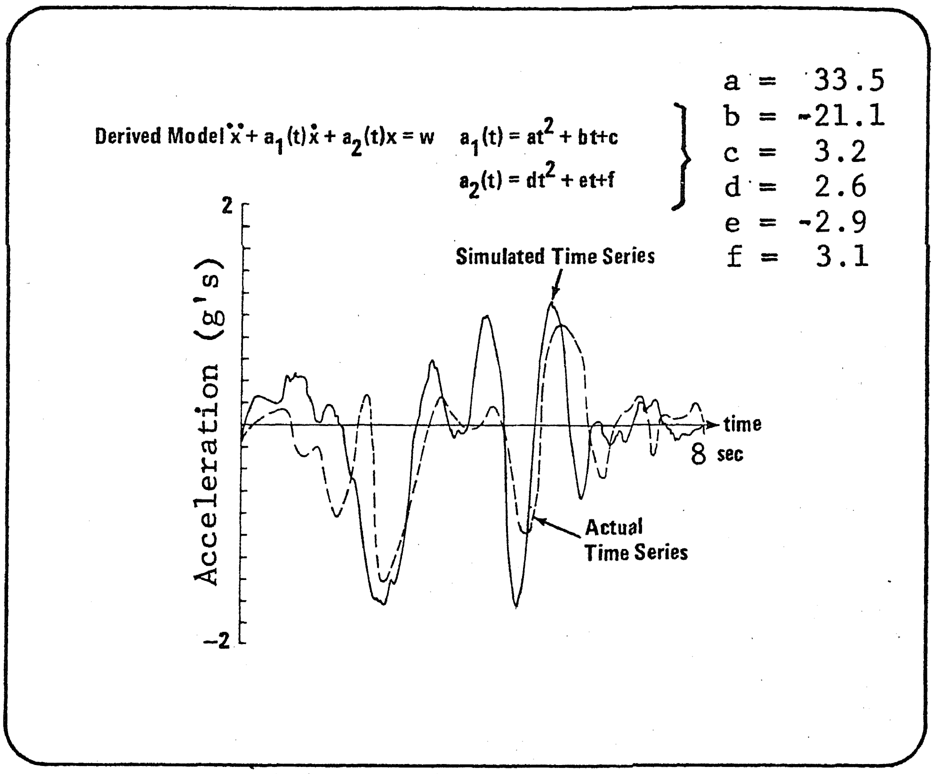

Fig. I El Centro South Earthquake - Comparison of Simulated and Actual Record

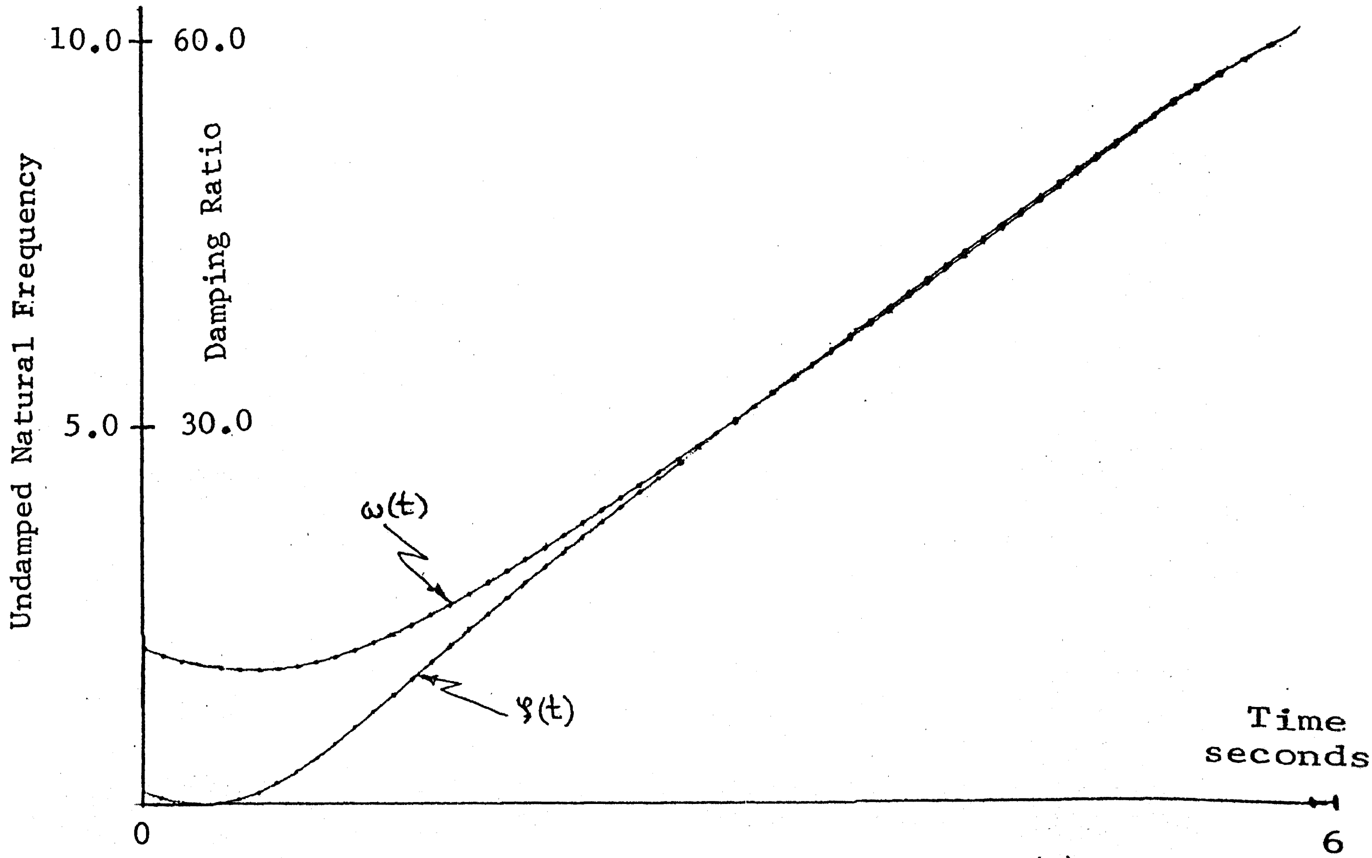

Fig. II Undamped Natural Frequency (ω) and Damping Ratio (ζ) vs. Time