AN APPROACH TO SIMULATE A LARGE SET OF MULTICORRELATED RANDOM PROCESSES

by

Tatsuo OHMACHI[I] and Yoshiaki KIMURA[II]

SUMMARY

This paper intends to make practical use of the methods previously proposed by many researchers for the simulation of multicorrelated random processes. The formal applications of those methods result in an extraordinarily large computational job with the increase in the number of dimensions or components involved in the processes. The procedure proposed here is based on the factor analysis concept which serves to construct a large number of processes by a smaller number of specific factors. Prior to the simulation, it is necessary to evaluate the factor loading matrix and factor scores. Instead of the original geometric processes, a series of derived factor vector is simulated to give the processes by superpostion.

INTRODUCTION

In an earthquake resistant design of a structure, it is usually assumed that the same motion acts simultaneously at all parts of the structure's foundation. This assumption is only satisfied if the base dimensions are small relative to the propagating seismic wave length and the foundation soil or rock is rigid enough for the rotational motion to be negligible. Thus, for a case in which the relative movements are expected at different points of the foundation, it is more appropriate to apply different earthquake inputs at separate points of supports. This is well known as multiple support excitation.

Due to the lack of simultaneous earthquake records available for the seismic design and the knowledge on such ground motion characteristics, a number of methods to produce artificial earthquake motions have been proposed, some stationary and others nonstationary. However, the formal application of these methods is limited to cases in which the number of components or dimensions is rather small. This is because the evaluation of the mutual relationships between each process followed by the simulation will result in an exponentially tedious calculation with the increase in the number of dimensions concerned.

To make practical use of those methods for the simulation of a large set of multicorrelated random processes, one of the measures is to reduce the number of dimensions in an effective manner. This implies that every component should be simulated in accordance with the significance of the intercorrelations as well as that of the effects on the structure. In other words, a simulation of a large set of such processes will be accomplished most efficiently by establishing firstly the significance of each component involved and then reducing the number of dimensions before performing the simulation.

[I] Associate Professor of Environmental Engineering, Tokyo Institute of Technology, Japan

[II] System Engineer, The Kaihatsu Computing Service Center, Ltd., Japan.

Factor analysis which is one of the fields of multivariate analysis has been developed for detecting informations latent in a large number of sample random variables and constructing them by smaller number of specific factors. Although the former development is not so great for the area of time series, the concept might be applied to various kinds of problems involved in earthquake engineering. This paper describes an approach in which the stationary multicorrelated random processes will be expressed in a form of the factor analysis model and the simulation procedure will be proposed.

A FACTOR ANALYSIS MODEL FOR MULTICORRELATED RANDOM PROCESSES

The fundamental concept of factor analysis is to express random variables $x_1, x_2, \ldots, x_p$ in terms of the common factors $f_1, f_2, \ldots, f_q$ and the residuals $e_1, e_2, \ldots, e_p$ by linear relations

$$\begin{aligned} x_{1m} &= a_{11}f_{1m} + a_{12}f_{2m} + \ldots\ldots + a_{1q}f_{qm} + e_{1m} \\ x_{2m} &= a_{21}f_{2m} + a_{22}f_{2m} + \ldots\ldots + a_{2q}f_{qm} + e_{2m} \\ &\vdots \\ x_{pm} &= a_{p1}f_{pm} + a_{p2}f_{2m} + \ldots\ldots + a_{pq}f_{qm} + e_{pm} \\ &(m=1,2,\ldots,n) \end{aligned} \tag{1}$$

in which the distribution of each x_m is assumed to be Gaussian with zero mean. The number of common factors q is usually less than that of sampled variables p , that is $q < p$. Alternatively Eq.(1) will be expressed in matrix form

$$\underset{(p\times n)}{X} = \underset{(p\times q)}{A} \cdot \underset{(q\times n)}{F} + \underset{(p\times n)}{E} \tag{2.a}$$

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & & & \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{bmatrix}, \quad F = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_q \end{bmatrix}, \quad E = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{bmatrix} \tag{2.b}$$

in which A is called the factor loading matrix or the factor patterns and e_i the unique factor. Eq.(2) is a simple statement that a set of observed variables can be made up of linear combinations of smaller number of common factors and the specific unique factors which include various kinds of error.

In general, common factors need not be orthogonal, but for simplicity in this paper they are assumed to be orthonormal and each unique factor is not correlated with other factors, that is,

$$E[\ F\ F^T] = I \quad \text{(identity matrix)} \tag{3.a}$$

$$E[\ e_i f_k^T] = 0 \tag{3.b}$$

$$E[\ e_i e_j^T] = 0 \qquad (i \neq j) \tag{3.c}$$

$$i,j=1,2,\ldots,p \qquad k=1,2,\ldots,q$$

where $E[\cdot]$ denotes the ensemble average and $(\cdot)^T$ the transpose. If each x_i

is normalized to have unit variance, the variance and the covariance of X and C (=AF) will be

$$E[\ X\ X^T] = R \tag{4.a}$$

$$E[\ C\ C^T] = E[\ A\ F\ F^T\ A^T] = A\ A^T \tag{4.b}$$

$$\begin{aligned} E[\ C\ X^T] &= E[\ A\ F\ (A\ F + E)^T] \\ &= A\ A^T = R - B \end{aligned} \tag{4.c}$$

in which R is the correlation matrix and B is a diagonal matrix named the uniqueness matrix. The element of B is given as

$$b_{ii} = 1 - \sum_{k=1}^{q} (a_{ik})^2 \tag{5}$$

It is convenient to introduce the communality h_i and the uniqueness b_i defined by

$$h_i = \sum_{k=1}^{q} (a_{ik})^2, \quad b_i = 1 - h_i \tag{6}$$

From Eqs.(4) through (6), it is evident that the communality means the contribution of the portion composed of common factors to the whole variation. With high communality, the variation involved in a specific variable can be explained to an adequate extent by the common factors, while with low communality, in other words, with high uniqueness, the contribution of the unique factor to the whole variation is very significant.

The factor analysis concept mentioned above would be directly applied to stationary multicorrelated random processes. Replacing the index m by time t in Eq.(1) leads to

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_p(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & & & \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{bmatrix} \begin{bmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_q(t) \end{bmatrix} + \begin{bmatrix} e_1(t) \\ e_2(t) \\ \vdots \\ e_p(t) \end{bmatrix} \tag{7}$$

For convenience, Eq.(7) will be expressed symbolically as

$$x(t) = A\ f(t) + e(t) \tag{8}$$

The relations in Eqs.(3) through (6) will be similarly introduced for Eq.(8).

EVALUATION OF FACTOR LOADING MATRIX AND FACTOR SCORES

A number of techniques have been developed to estimate a factor loading matrix A from a correlation matrix R . Among them, the employed in this paper is the canonical factor analysis procedure, which will be briefly described here with an example. In a word, this procedure determines A in such a way that the observed variables x(t) and the portion represented by common factors C(t) (=Af(t)) will exhibit the maximum canonical correlation.

Consider two arbitrary linear combinations expressed as $u = y^T C(t)$ and $v = z^T x(t)$. Normalizing u and v to have unit variance, the algebraic problem is to find y and z to maximize the following equation:

$$\begin{aligned} \phi = E[\ y^T C(t) \cdot z^T\ x(t)] &- \frac{1}{2}\lambda\,(E[\ y^T C(t)]^2 - 1) \\ &- \frac{1}{2}\mu\,(E[\ z^T x(t)]^2 - 1) \end{aligned} \tag{9}$$

in which λ and μ are Lagrange multipliers. Differenciating Eq.(9) with respect to y and z , and equating to zero yield to an eigen problem as

$$[\, B^{-\frac{1}{2}}(R - B) B^{-\frac{1}{2}} - \theta_i^2]\, g_i = 0 \tag{10}$$

where

$$\theta_i^2 = \lambda_i^2/(1 - \lambda_i^2)\ ,\ g_i = B^{\frac{1}{2}}y \tag{11}$$

Solving Eq.(10) relative to eigen values θ_i^2 and the corresponding eigen vectors g_i provides the estimated A as follows.

$$A = B^{\frac{1}{2}} G\, \Theta \tag{12}$$

in which Θ is a (q×q) diagonal matrix having θ's as the components, and G is a (p×q) matrix made up by the corresponding eigen vectors.

Since A and B are as yet unknown at the beginning of the calculation, iterative methods are introduced in solving Eq.(10). To initiate the process, the initial B is usually assumed as

$$B = (\mathrm{dia.}(\ R^{-1}))^{-1} \tag{13}$$

By repeating the iteration shown in Fig.3, the factor loading matrix can be improved to any desired level of accuracy.

<u>Example.</u> To illustrate the concept described above, the factor analysis method was applied to the simultaneous records shown in Fig.1. After the respective waveform has been normalized, the correlation (covariance) matrix was calculated as in Tab.1. Tab.1 shows that the correlation between ⓑ and ⓓ is very low, comparing with other correlations.

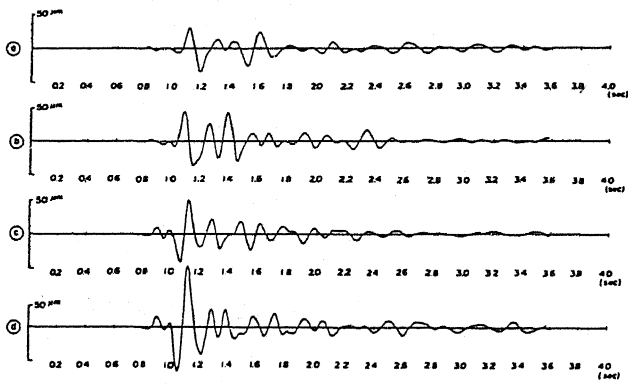

Fig.1 Original simultaneous records

	ⓐ	ⓑ	ⓒ	ⓓ
ⓐ	1.000	0.221	0.373	0.428
ⓑ	0.221	1.000	-0.242	0.051
ⓒ	0.373	-0.242	1.000	0.624
ⓓ	0.428	0.051	0.624	1.000

Tab.1 Correlation matrix

	(1)	(2)	(3)	(4)
θ^2	5.534	1.792	0.239	-0.000
λ	0.920	0.801	0.439	---

Tab.2 Eigenvalues and canonical correlation coefficients

	(1)	(2)	(3)
ⓐ	0.508	0.403	0.260
ⓑ	-0.126	0.735	-0.040
ⓒ	0.855	-0.180	0.042
ⓓ	0.780	0.194	-0.176

Tab.3 Factor loading matrix

The canonical correlation coefficients λ corresponding to the eigenvalues θ^2, and the factor loading matrix A are shown in Tabs.2 and 3, respectively. From Tab.2, it may be pointed out that the first two loading factors shown columnwise in Tab.3 are highly significant. Fig.2 shows the communality and the uniqueness, which means that almost all the variation involved in the waveform of ⓒ and ⓓ can be reduced to those of the factor (1) and the unique factor, while the variation involved in the waveform of ⓑ can be satisfactorily explained by the factor (2) and the unique factor.

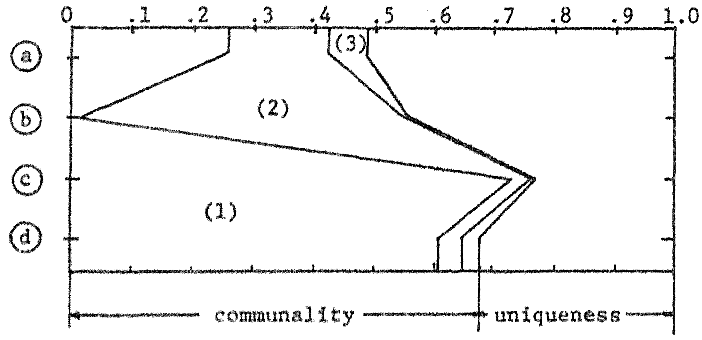

Fig.2 Communality and uniqueness

Once the factor loading matrix A has been obtained, the factor scores f(t) can be evaluated from the sample variables x(t) by the linear relations as

$$f(t) = W^T x(t) \tag{14}$$

in which the matrix W is a (p×q) weighting matrix. Imposing the orthonormal conditions on the estimated factor scores will provide the algebraic simplification in formulating the simulation procedure. The matrix given by

$$W = R^{-1}A(A^T R^{-1} A)^{-\frac{1}{2}} \tag{15}$$

is an example which makes the estimated factor scores orthonormal. In Eq.(15), the last term in the right-hand side represents

$$(A^T R^{-1} A)^{\frac{1}{2}} = P\Lambda^{\frac{1}{2}} P^T \tag{16}$$

in which Λ is a diagonal matrix composed of the eigenvalues of the matrix product $A^T R^{-1} A$ and the matrix P has the corresponding eigenvectors in columns. Now, the residual e(t) will be given by

$$e(t) = x(t) - Af(t) \tag{17}$$

SPECTRAL RELATIONSHIPS BETWEEN SAMPLE VARIABLES AND FACTORS

If the sample processes and the factors are in linear relationship as in Eq.(7), the cross-correlation function between the members of the samples will be

$$R_{x_{ij}}(\tau) = E[\,x_i(t)x_j(t+\tau)]$$

$$= E[\{\sum_k a_{ik}f_k(t) + e_i(t)\}\{\sum_l a_{jl}f_l(t+\tau) + e_j(t+\tau)\}]$$
$$= \sum_k\sum_l a_{ik}a_{jl}E[f_k(t)f_l(t+\tau)] + \sum_k a_{ik}E[f_k(t)e_j(t+\tau)]$$
$$+ \sum_l a_{jl}E[f_l(t+\tau)e_i(t)] + E[e_i(t)e_j(t+\tau)] \qquad (18)$$

As is evident from the definition, each uniqu factor represents the unique characteristics of the respective component as well as various kinds of error terms. Thus, for practical purposes it will be permissible to assume that each unique factor is statistically independent of other factors, that is,

$$E[f_k(t)e_j(t+\tau)] = 0 \qquad (19.a)$$
$$E[f_l(t+\tau)e_i(t)] = 0 \qquad (19.b)$$
$$E[e_i(t)e_j(t+\tau)] = \delta_{ij}R_{e_{ij}}(\tau) \qquad (19.c)$$

in which δ_{ij} is Kronecker's symbol. Substituting Eqs.(19) into Eq.(18) leads to

$$R_x(\tau) = A\ R_f(\tau)\ A^T + R_e(\tau) \qquad (20.a)$$

where

$$R_x(\tau)=\begin{bmatrix} R_{x_{11}}(\tau) & R_{x_{12}}(\tau) & \dots & R_{x_{1p}}(\tau) \\ R_{x_{21}}(\tau) & R_{x_{22}}(\tau) & \dots & R_{x_{2p}}(\tau) \\ \vdots & & & \\ R_{x_{p1}}(\tau) & R_{x_{p2}}(\tau) & \dots & R_{x_{pp}}(\tau) \end{bmatrix}, \quad R_f(\tau)=\begin{bmatrix} R_{f_{11}}(\tau) & R_{f_{12}}(\tau) & \dots & R_{f_{1q}}(\tau) \\ R_{f_{21}}(\tau) & R_{f_{22}}(\tau) & \dots & R_{f_{2q}}(\tau) \\ \vdots & & & \\ R_{f_{q1}}(\tau) & R_{f_{q2}}(\tau) & \dots & R_{f_{qq}}(\tau) \end{bmatrix}$$

$$R_e(\tau)=\begin{bmatrix} R_{e_{11}}(\tau) & 0 & \dots & 0 \\ 0 & R_{e_{22}}(\tau) & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & R_{e_{pp}}(\tau) \end{bmatrix} \qquad (20.b)$$

Consequently, the spectral density function for the sample processes will be given in the matrix form

$$S_x(\omega) = A\ S_f(\omega)\ A^T + S_e(\omega) \qquad (21.a)$$

where

$$S_x(\omega) = \begin{bmatrix} S_{x_{11}}(\omega) & S_{x_{12}}(\omega) & \dots & S_{x_{1p}}(\omega) \\ S_{x_{21}}(\omega) & S_{x_{22}}(\omega) & \dots & S_{x_{2p}}(\omega) \\ \vdots & & & \\ S_{x_{p1}}(\omega) & S_{x_{p2}}(\omega) & \dots & S_{x_{pp}}(\omega) \end{bmatrix}, \text{ etc.,} \qquad (21.b)$$

In Eqs.(20) and (21), it should be noted that the order of the cross-spectral matrices is (p×p) in the left-hand side, while (q×q) in the right-hand side.

Therefore, for the usual cases except when all the processes are highly correlated, the number of the cross-correlated terms will be remarkably reduced by use of a factor analysis model.

SIMULATION PROCEDURE BASED UPON FACTOR ANALYSIS CONCEPT

Since such quantities as the means, the variances and the spectral density functions of factor scores, have been given, the simulation of factor scores will be possible by employing some techniques already established. With simulated $f(t)$ and $e(t)$, the desired original geometric processes $x(t)$ will be built up by

$$\tilde{x}(t) = A\,\tilde{f}(t) + \tilde{e}(t) \qquad (22)$$

in which the tilde denotes the simulated values.

Fig. 3 summarizes the simulation procedure described in this paper. Evaluation of a factor loading matrix and factor scores is a process additional to the usual procedures but it will serve not only to reduce the following tedious calculations, but also for designers to understand the intercorrelations of the random processes.

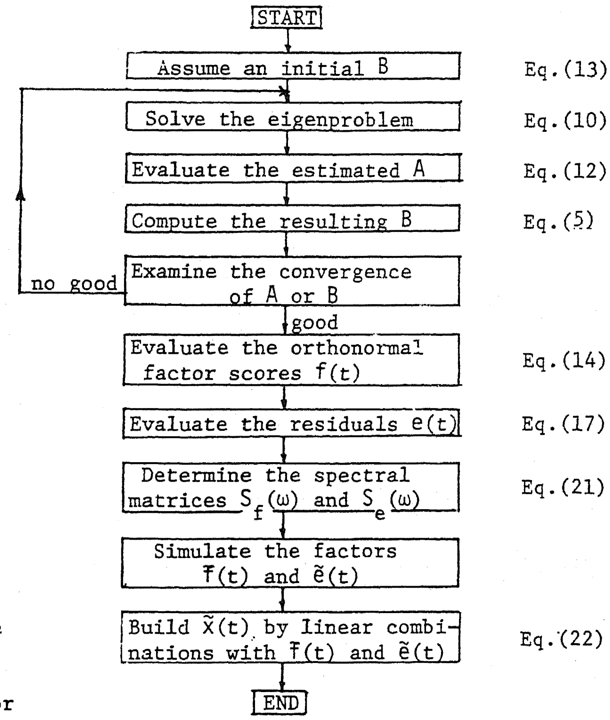

Fig.3 Summary of the procedure proposed in this paper

CONCLUDING REMARKS

Expressing multicorrelated random processes by the linear combinations of loading factors and factor scores gives an impression similar to representing arbitrary deflections of multi-degrees-of-freedom vibrating system by vibration mode shapes and the modal amplitudes. One of the differences lies in the fact that the evaluated loading matrix retains a kind of uncertainties relative to matrix rotation. Canonical factor analysis is known to give the most-likelihood estimation of the matrix and to provide the method to test the significance of the evaluated factor patterns.

The method proposed in this paper will be useful in simulating a large set of multicorrelated random processes which has been tedious enough for practical purposes, but some assumptions employed here need to be examined by a variety of actual earthquake records.

ACKNOWLEDGMENT

The authors appreciate Electric Power Development Co., Ltd., for partial support of this research.

REFERENCES

1. Anderson, T. W., "An Introduction to Multivariate Statistical Analysis," John Wiley & Sons, Inc., 1958.

2. Clough, R. W., and Penzien, J., "Dynamics of Structures," McGraw-Hill Kogakusha, Ltd., 1975.

3. Parzen, E., "Stochastic Processes," Holden-Day, Inc., 1967.

4. Hoshiya, M., "Simulation of Multi-Correlated Random Processes and Application to Structural Vibration Problems," Proc. JSCE, No. 204, 1972, pp. 121-128.

5. Ishii, K., "The Correlated Nonstationary Random Processes for Earthquake Inputs," Proc. JSCE, No. 259, 1977, pp. 41-54.

6. Shinozuka, M., "Digital Simulation of Ground Accelerations," Proc. 5WCEE, 1974, pp. 2829-2838.

7. Hoshiya, M. and Isoyama, R., "Simulation of Multi-Dimensional Nonstationary Earthquake Accelerations," Proc. JSCE, No. 269, 1978, pp. 41-52

8. Kubo, T. and Suzuki, N., "An Application of Synthetic Earthquake Ground Motions to Response Analysis - Simulation of Earthquake Ground Motions," Proc. 5th Japan Earthq. Engng Symp., 1978, pp. 89-96.

9. Kubo, T. and Penzien, J., "Analysis of Three-Dimensional Earthquake Ground Motion along an Orthogonal Set of Principal Axes," Proc. 5JEES, 1978, pp. 97-104.

10. Watabe, M. and Tohdo, M., "Reaearch on the Simulation of Three-Dimensional Earthquake Ground Motions," Proc. 5JEES, 1978, pp. 105-120.