STOCHASTIC INTERPOLATION OF POWER SPECTRA
USING OBSERVED EARTHQUAKE GROUND MOTIONS

H. MORIKAWA* and H. KAMEDA**

*School of Civil Engineering, Kyoto University,
Yoshida-Honmachi, Sakyoku, Kyoto 606-01, Japan
**Disaster Prevention Research Institute, Kyoto University,
Gokasho, Uji, Kyoto 611, Japan

ABSTRACT

Methods are presented for stochastic interpolation of power spectra using observed earthquake ground motion time histories. For this purpose, we extend the theory of conditional random fields developed by Kameda and Morikawa (1994). Then properties of spectral uncertainties are made clear and we derive the estimation errors of interpolated spectra. Furthermore, in order to examine the applicability of the presented methods, numerical examples are presented using simulated earthquake ground motions.

KEYWORD

Conditional random fields; stochastic interpolation; earthquake ground motion; frequency series; homogeneous random fields; weakly inhomogeneous random fields; power spectrum; quefrency domain; cepstrum.

INTRODUCTION

The goal of this study is establishment of methodology for the stochastic interpolation of earthquake ground motion fields. For a step of this purpose, we develop the theoretical framework to identify the spectral properties using the observed time series.

The theories or methodology with respect to the random fields involving the observed data as condition, which we call "conditional random fields" (CRF), have been investigated from various viewpoints by many researchers (e.g. Kameda and Morikawa, 1994; Kawakami, 1989; Vanmarcke et al., 1991). Most of these methodology, however, require a priori information about the spectral properties of the random fields. According to circumstance, these requirements in the formulation may restrict the applicability of the methodology to actual phenomena.

If we estimate a random field using a certain 'fixed' spectral properties determined on the basis of the obscure knowledge, we must be prepared for risk that the 'fixed' ones are methodically different from 'true'; namely, that we may base the estimation which is quite different from true one. Moreover, there is no way to make such differences clear.
Therefore, we propose two methods to estimate spectra with uncertainties on the basis of the observed time series. The uncertainties are caused by various sources of randomness: lack of information, microscopic fluctuation of source and medium, and so on. Hereafter, we call such type of spectra 'stochastic spectra' in order to distinguish from 'fixed' ones. In the following sections, we discuss the stochastic spectra with earthquake ground motions in mind, however, presented methods are useful for general time-space random fields such as wind speed fluctuations, sea waves, etc.

**STOCHASTIC INTERPOLATION OF GROUND MOTION SPECTRA**

*Discussion of Stochastic Spectra using Cepstrum Concepts*

In order to deal with power spectra as stochastic spectra, it is necessary to discuss stochastic processes on frequency domain, so-called "frequency series." Generally, the stochastic properties of frequency series are provided by cepstra defined in quefrency domain. In addition to the cepstra, we must introduce the "cross cepstra" in the quefrency domain to treat multi-variate frequency series which we shortly bring up. Thus, we are going to discuss both the time and frequency series freely using parameters defined in three domains of time, frequency and quefrency domain with the Fourier transformation and the inverse one.

To precisely discriminate among parameters belonging in the each domain, we use superscripts \((t), (f),\) and \((q)\) for time, frequency, and quefrency domain, respectively, if necessary. Particularly, since there are no terms for coherence and phase lag functions derived from cross cepstra in quefrency domain to the best of the authors' knowledge, the domain in which these functions belong is specified by the superscript \((f)\) or \((q)\) attached on the right shoulder.

In dealing with the stochastic spectra, we must surmount some difficulties. Many of them regarding the way to give the shape of cepstra. However, under the consideration that the uncertainties of spectra are caused by sources of randomness mentioned in the previous section, we are safe in treating the fields of stochastic spectra as homogeneous random fields except for the peculiar vibration character effected by distinct or deterministic variations of medium or ground structure. In other words, if there is little structural irregularity significantly affecting the general features of the vibration character, we consider that cepstra are the same at all the sites and that coherence functions\(^{(q)}\) between two frequency series are modeled in the following function which is the form often used for homogeneous fields (Chernov, 1960):

\[
Coh^{(q)}(x_0, q) = \exp[-\bar{a} x_0 q],
\]

where \(\bar{a}\) is a constant, \(q\) is the quefrency and \(x_0\) is distance between the two sites studying the frequency series. The phase lag functions\(^{(q)}\) are set to zero for any combination of a pair of sites because it is unnatural to consider the situation that the spatial transition of peaks or dips appearing on power spectra is orderly in the case where the sites are located on ground with the similar structure.

**Method I — for homogeneous ground condition**  "Method I" provides a basic methodology to utilize for the limited case where time series are observed at some sites on stochastically homogeneous ground conditions. This method has basis in the fact that if the time series \(U_i(t)\) at site \(i\) \((i=1,2, \ldots, n)\) is a zero-mean Gaussian process, its discrete Fourier coefficients series \(A_{ik}\) and \(B_{ik}\) are also zero-mean Gaussian processes, where \(U_i(t)\) is stochastic process. In this section, we show the way to stochastically interpolate the power spectra, following the framework of Method I established by Morikawa and Kameda(1994).

Replacing time series \(U_i(t)\) with frequency series \(A_{ik} = A_i(f_k)\) or \(B_{ik} = B_i(f_k)\) \((i = 1,2, \ldots, n)\) and introducing
quefrency \( q \), we obtain the Fourier series in the quefrency domain, as follows:

\[
A_i(f) = \sum_k \left\{ \hat{A}_{ik} \cos(2\pi q_k f) + \hat{B}_{ik} \sin(2\pi q_k f) \right\} \\
B_i(f) = \sum_k \left\{ \hat{A}_{ik} \cos(2\pi q_k f) + \hat{B}_{ik} \sin(2\pi q_k f) \right\}.
\]  

For our goal, hereafter, we discuss Eq.(2) applying the theory of conditional random fields (Kameda and Morikawa, 1994) to frequency-quefrency domain.

Then we statistically determine the cepstra, which describe the stochastic properties of the frequency series, by observed time series. On the basis of the supposition that the random fields\( f \) are homogeneous, the arithmetic means of observed cepstra, which are calculated from observed time series, are used as cepstra at all the sites. For the coherence function\( \omega \), we use the Eq.(1) determining the value of parameter \( \omega \) with the method of least squares. From the above procedure, we can immediately reach the stochastic properties of the conditioned Fourier coefficient series \( A_j(f \mid \text{cnd.}) \), \( B_j(f \mid \text{cnd.}) \) and simulate the realized values of Fourier coefficients at site \( j \) \( (j = m + 1, \ldots, n) \); where the word “cnd.” means that the quantity under discussion is conditioned by \( u_i(t); i = 1, 2, \ldots, m \) which are realized time series of \( U_i(t) \) and \( m \) is the number of observation sites. Furthermore, the power spectra are obtained through smoothed values of “raw” power spectra \( (A_n(f \mid \text{cnd.})^2 + B_n(f \mid \text{cnd.})^2)/4 \).

There are theoretically few difficulties in estimating the Fourier coefficient series\( f \) conditioned by observed data. In consideration of the power spectra, which is our goal, it is to be noted that the power spectra contain no information with regard to phase angles except for minimum phases, while the Fourier coefficients\( f \) do contain the information. In other words, the power spectra are independent of the phase angles, while the Fourier coefficients are dependent upon them. This means that Method 1 derives the various power spectra as a function of an apparent velocity \( = \text{phase lags} \) for the same dataset of observed time series.

In order to avoid such problems regarding the phase angles, we must correct the phase lags between the observed time series before we set about the stochastic interpolation of power spectra. Thus, calculating the phase lag functions\( f \phi_{ip}(f) \) between the sites 1 and \( p \) \( (p = 2, 3, \ldots, m) \), we shift the phase angles of \( U_p(t) \) for each harmonic components so as to reduce \( \phi_{ip}(f) \) to zero, which is the technique usually used.

**Method II — for weakly inhomogeneous ground conditions** Since Method I is theoretically simple, we can readily reach the goal, which is the stochastic interpolation of power spectra, using the theory of conditional random fields. However, the applicability of Method I to real phenomena is limited by contrary factors. Although we assume that the time series are observed at sites with homogeneous ground structures, it is difficult to carry out the observation of earthquake ground motions under such ideal conditions. In this section, in order to achieve more extensive applicability we devise Method II in which we can use records observed at the sites with weakly inhomogeneous ground structures as the condition.

In this method, in order to consider the variety of ground structures, we divide the stochastic spectra into the term reflecting the deterministic feature of the ground structure and randomness from the deterministic term. Thus, the power spectrum \( S_{ii}(f) \) at site \( i \) \( (i = 1, 2, \ldots, n) \) are represented by

\[
S_{ii}(f) = S_{ii}^D(f)S_{ii}^R(f),
\]

where \( S_{ii}^D(f) \) denotes the deterministic site effect which is independent of events, and \( S_{ii}^R(f) \) is the random component with different stochastic properties for every event. Since the specific site characteristics can be represented by the relative ratio to reference site, as which we set site 1 without loss of generality, Eq.(3) is rewritten as

\[
S_{ii}(f) = S_{ii}^D(f)S_{ii}^R(f) \quad (i = 1, 2, \ldots, n),
\]

where \( S_{ii}^R(f) \) is the relative ratio to the reference site.
where $\mathcal{S}_u^f(f) = \mathcal{S}_u^R(f)/\mathcal{S}_u^I(f)$. To apply Method I to $\mathcal{S}_u^R(f)$ in Eq.(3) (or $\mathcal{S}_u^I(f)$ in Eq.(4)), we need to extract the contribution to $\mathcal{S}_u^R(f)$ and $\mathcal{S}_u^I(f)$ in Eq.(3) (or $\mathcal{S}_u^R(f)$ and $\mathcal{S}_u^I(f)$ in Eq.(4)) from Fourier coefficients series $A_k(f)$ and $B_k(f)$ in spite of the essential difficulties. Therefore, we consider that it is appropriate to deal with power spectra instead of Fourier coefficients in order to separate the power spectrum $\mathcal{S}_u(f)$ into the deterministic and the random components.

If $U_i(t)$ is a stationary Gaussian process with zero-mean, its "raw" power spectrum follows the $\chi^2$ distribution with a degree of freedom 2. However, since the power spectrum $\mathcal{S}_u(f)$ is ordinarily estimated in smoothed value of the "raw" power spectrum, its probability distribution is $\chi^2$ distribution with large freedom. In this study, to treat power spectra as the stochastic processes in the frequency domain substitute the log-normal distributions for theirs. The marginal PDF of power spectrum $\mathcal{S}_{iik}$ at frequency $f_k$ and site $i$ are represented as

$$f_{\mathcal{S}_{iik}}(s_{iik}) = \frac{1}{\sqrt{2\pi}\sigma_{iik}} \exp \left[ -\frac{1}{2} \left( \frac{\ln s_{iik} - \lambda_{iik}}{\sigma_{iik}} \right)^2 \right] \quad (i = 1, 2, \ldots, n).$$

(5)

There is little problem in this appropriation because $\chi^2$ distributions have large freedom and we are not after accuracy of the probability of extreme values. It is convenient, if anything, to use the log-normal distribution for Eqs.(3) and (4) decomposed into product. Thus, we discuss the stationary Gaussian processes derived from the logarithm of both sides of Eqs.(3) and (4):

$$\ln \mathcal{S}_u(f) = \ln \mathcal{S}_u^R(f) + \ln \mathcal{S}_u^I(f), \quad \ln \mathcal{S}_u(f) = \ln \mathcal{S}_u^R(f) + \ln \mathcal{S}_u^I(f) \quad (i = 1, 2, \ldots, n).$$

(6)

We determine $\mathcal{S}_{jji}^g(f)$ (or $\ln \mathcal{S}_{jji}^g(f)$) at site $j$ ($j = m + 1, \ldots, n$), where we have no information regarding the ground motion, using the linear interpolation of observed values at site $i$ ($i = 1, 2, \ldots, m$) because of the supposition that $\mathcal{S}_{jji}^g(f)$ (or $\mathcal{S}_{jji}^g(f)$) are deterministic values changing gradually in the study area.

We can use only $\mathcal{S}_{jji}^g(f)$ in the case where we must estimate the site effects on the basis of the observed values without a priori information with respect to them. We show the method of determination of $\mathcal{S}_{jji}^g(f)$ using the observed values. From the assumption that $\mathcal{S}_{jji}^g(f)$ belongs on a homogeneous random field, namely $\mathcal{S}_{jji}^g(f)$ has one and the same population at all the sites, the arithmetic mean of $\mathcal{S}_{jji}^g(f)$ over many events may be independent of the site. Therefore, we can obtain the following relation:

$$\mathcal{S}_{jji}^g(f) = \frac{\mathcal{S}_{jji}^g(f)}{\mathcal{S}_{jji}^R(f)} = \frac{\mathcal{S}_{jji}^g(f)}{\mathcal{S}_{jji}^R(f)} \cdot \frac{E_a[\mathcal{S}_{jji}^g(f)]}{E_a[\mathcal{S}_{jji}^g(f)]} = \frac{E_a[\mathcal{S}_{jji}^g(f)]}{E_a[\mathcal{S}_{jji}^g(f)]},$$

(7)

where $E_a[\cdot]$ denotes the arithmetic mean over events. Since $E_a[\mathcal{S}_{jji}^g(f)]/E_a[\mathcal{S}_{jji}^g(f)]$ can be replaced by the geometrical mean of $\mathcal{S}_{jji}^g(f)/\mathcal{S}_{jji}^R(f)$, the optimal estimator of $\mathcal{S}_{jji}^g(f)$ is updated every event using the geometrical mean over all the spectral ratios $\mathcal{S}_{jji}^g(f)/\mathcal{S}_{jji}^R(f)$ observed in the study event and former events.

If we have a priori information with regard to $\mathcal{S}_{jji}^g(f)$ (or $\mathcal{S}_{jji}^g(f)$), we can obtain the optimal estimator of $\mathcal{S}_{jji}^g(f)$ (or $\mathcal{S}_{jji}^g(f)$) updating it every event by means of the method of Bayes' statistics. This means that we can exhaustively use the information we have using the presented method.

On the other hand, the frequency series $\ln \mathcal{S}_{jji}^g(f)$ ($i = 1, 2, \ldots, n$) are stationary Gaussian processes with zero-mean. As $\mathcal{S}_{jji}^g(f)$ represents the random component excluding site effects, the cepstra and cross cepstra of $\ln \mathcal{S}_{jji}^g(f)$ can be determined under the same treatments mentioned in previous section. As similar to the Fourier coefficient series, we represent in $\ln \mathcal{S}_{jji}^g(f)$ as

$$\ln \mathcal{S}_{jji}^g(f) = \sum_k \left\{ A_{ik} \cos(2\pi q_k f) + B_{ik} \sin(2\pi q_k f) \right\},$$

(8)

and we can immediately obtain the stochastic characteristics of $\ln \mathcal{S}_{jji}^g(f)$ ($j = m + 1, \ldots, n$) conditioned by the observed time series using the theory of conditional random fields.
The \( \ln S_{ij}^\mathcal{F}(f) \) may be non-zero-mean Gaussian process because \( S_{ii}^\mathcal{F}(f) \), which denotes a fluctuation of the spectral ratio to the spectrum observed at site \( i \), includes the contribution of the deterministic term \( S_{ii}^\mathcal{D}(f) \). Therefore, dividing \( \ln S_{ii}^\mathcal{F}(f) \) into the DC-component and the remainder, which is the stationary Gaussian process with zero-mean, we make their interpolation separately and then combine them. On one hand, the DC-component of \( \ln S_{ij}^\mathcal{F}(f) \) at site \( j \) \( (j = m + 1, \ldots, n) \) is obtained by means of the linear interpolation as the case of \( S_{ii}^\mathcal{D}(f) \) (or \( S_{ij}^\mathcal{D}(f) \)). On the other hand, the remaining zero-mean Gaussian process is stochastically interpolated as the case of \( \ln S_{ij}^\mathcal{F}(f) \). Finally, we obtain \( \ln S_{ij}^\mathcal{F}(f) \) from the sum of the DC- and the random components.

From the above procedure, even though we have no a priori information regarding the spectral characteristics, we can determine the site effect \( S_{ij}^\mathcal{F}(f) \) and the random component \( S_{ij}^\mathcal{F}(f) \) at site \( j \) \( (j = m + 1, \ldots, n) \); namely, we use the linear interpolation and the update by means of the geometrical mean of spectral ratio for \( S_{ij}^\mathcal{D}(f) \), and the linear interpolation and the theory of conditional random fields for \( S_{ij}^\mathcal{F}(f) \).

For the determination of \( S_{ij}(f) \) \( (j = m + 1, \ldots, n) \), we derive \( \lambda_{jk} \) and \( \zeta_{jk} \) in Eq.(5) combining \( S_{ij}^\mathcal{F}(f) \) and \( S_{ij}^\mathcal{D}(f) \) as follows:

\[
\lambda_{jk} = \ln S_{ij}^\mathcal{F}(f_k) + E[\ln S_{ij}^\mathcal{F}(f_k) | \text{cnd.}], \quad \zeta_{jk} = \text{Var}[\ln S_{ij}^\mathcal{F}(f_k) | \text{cnd.}] \quad (j = m + 1, m + 2, \ldots, n), \tag{9}
\]

where \( S_{ij}^\mathcal{F}(f_k) \) is a estimated value of \( S_{ij}^\mathcal{F}(f) \). Eq.(9) provides the stochastic character of the power spectrum \( S_{ij}(f) \) at site \( j \) conditioned by \( S_{ii}(f) \) \( (i = 1, 2, \ldots, m) \) under the approximations of Eq.(5).

Note the following issues: in the case where there is no a priori information with respect to the power spectra, since we cannot detect the random component, \( S_{ii}(f) \) becomes \( S_{ii}(f) \) \( (i = 1, 2, \ldots, m) \) for the first event. Therefore, \( S_{ij}(f) \) \( (j = m + 1, \ldots, n) \) is determined by means of the linear interpolation of \( \ln S_{ii}^\mathcal{F}(f) \) \( (i = 1, 2, \ldots, m) \) as the earthquake ground motions are observed for the first time.

### Table 1 Specifications of pseudo-earthquakes

<table>
<thead>
<tr>
<th>Magnitude</th>
<th>Epicentral Dis.</th>
<th>Apparent Veloc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ1</td>
<td>6.0</td>
<td>80</td>
</tr>
<tr>
<td>EQ2</td>
<td>7.5</td>
<td>90</td>
</tr>
<tr>
<td>EQ3</td>
<td>6.0</td>
<td>40</td>
</tr>
<tr>
<td>EQ4</td>
<td>7.0</td>
<td>150</td>
</tr>
<tr>
<td>EQ5</td>
<td>6.0</td>
<td>100</td>
</tr>
<tr>
<td>EQ6</td>
<td>6.5</td>
<td>60</td>
</tr>
<tr>
<td>EQ7</td>
<td>8.0</td>
<td>200</td>
</tr>
</tbody>
</table>

**Comparison of the Two Methods**

In order to compare the two methods and examine their applicability, we make a “blind test” using random fields which are numerically generated under a given stochastic properties. In the blind test, we simulate a random field in time-space as a target and compare the power spectra of target with the ones estimated by means of Method I or II under the condition that part of the target is observed.
We generate a pseudo-earthquake ground motion field, which is used as the target, at five sites on the $z$ axis shown in Fig. 1 using the technique of conditional simulation (Kameda and Morikawa, 1994). Then, we give appropriately seven sets of magnitude $M$, epicentral distance $\Delta$, and apparent velocity $v$ as shown in Table 1 as pseudo-earthquake, and one-sided power spectra at site $S_p$ and cross spectra between sites $S_p$ and $S_q$ ($p, q = 1, \ldots, 5; p < q$) as follows (Kameda, 1987; Kawakami and Sato, 1988):

\[
\begin{align*}
C_r(f) &= \gamma^2 \frac{2\beta_p}{\pi^2 f_0} \left\{ 1 - \left( \frac{f}{f_0} \right)^2 \right\}^2 \times r(f, z_p), \\
S_{pq}(f, z_0) &= \sqrt{S_{pp}(f)S_{qq}(f)} \exp \left[ -\frac{ifz_0}{v} \right] \exp \left[ -\alpha \frac{|f| z_0}{2v} \right],
\end{align*}
\]

where $i = \sqrt{-1}$, $\gamma$, $f_0$, and $\beta_p$ are respectively peak RMS acceleration, predominant frequency, and band width, which are regression function of $M$ and $\Delta$ in km, and $\alpha$ is distortion constant which is fixed at 1.0 in this study. $r(f, z_p)$ represents the spectral ratio between the sites $S_1$ and $S_p$. In following calculation, we consider the two cases:

- **Case 1:** The power spectra are same at all the sites; $r(r, z_p) \equiv 1$,

- **Case 2:** The power spectra have some differences;

\[
r(f, z_p) = 1.0 + \frac{\frac{1}{\sqrt{10\pi}} \left( \frac{f}{\sqrt{350\pi}} \right)^2}{\left\{ 1 - \left( \frac{f}{\sqrt{350\pi}} \right)^2 \right\}^2 + \frac{1}{10\pi} \left( \frac{f}{\sqrt{350\pi}} \right)^2}.
\]

Note that the author does not intend to claim these formulas Eq.(10) as a standard form of the ground motion spectra. As a purpose of this study is the construct a probabilistic framework of the subject matter, they are used just for the purpose of numerical illustration. Moreover, above equation forms are not important, because our methods do not require such a priori information and are independent of the particular model representing the earthquake ground motion.

We regard a wave field obtained through the above procedure as the target. The target field is compared with the simulated power spectra at sites $S_2$, $S_3$, and $S_4$ by Method I or II under the condition that the wave forms are observed at sites $S_1$ and $S_5$. Of course, we have no information about the "true" spectral characteristics.
Fig. 3 Comparison between simulated power spectra and corresponding targets for Case 2 (left column: observed power spectra, middle column: simulated by means of Method I, right column: simulated by means of Method II)

Fig. 4 Conditional mean values and standard deviation of the power spectra estimated by using Method II (EQ7, Case 2)

Figs. 2 and 3 show the simulated power spectra and corresponding target ones for EQ1 and EQ7 under Cases 1 and 2, respectively, using the two methods, where the results for EQ1 by Method II are obtained through the linear interpolation. In Case 1 (Fig. 2), we cannot point out the obvious difference in EQ7 between the results obtained through the two methods. On the other hand, in Case 2 (Fig. 3), because the different spectra are given at each site, there are some marked differences between the target and the simulated spectrum. Especially the target and the spectra simulated by Method I are sometimes completely different because of the inhomogeneous fields such as Case 2 in the method. However, the results from Method II are more harmonious with the target than Method I.

These numerical results confirm our expectation that Method II is more applicable to various ground conditions than Method I. Therefore, in the following analysis, we discuss Method II without notice.

**VARIANCES OF INTERPOLATED SPECTRA**

It is easy to derive the "raw" power spectra simulated by Method I, which are the spectra before the smoothing and follow the $\chi^2$ distribution. The estimation error of smoothed power spectra, however, depends on the
method of smoothing. Thus we must consider the various methods of smoothing and analyze the estimation error for each method. On the other hand, for Method II we can uniquely derive the estimation error using the parameters for log-normal distribution represented in Eq.(9). In Method II, it should be noted that the systematic errors are involved with the approximation of the probabilistic distribution, however, we consider that such error is not important for our goal as we have mentioned in the previous section.

Fig.-4 shows the conditional mean values and variances of the power spectra estimated by using Method II in case where the wave forms are observed at two or three observation sites for the pseudo earthquake fields of Case 2. It is observed that the differences between the target and simulation are small with increasing the number of observation sites and decreasing the distance between the observation site and the site where the power spectrum is estimated.

CONCLUSIONS

On the basis of the theory of conditional random fields, we established two methods for stochastic interpolation of power spectra using the observed time series of earthquake ground motion. While in one method (Method I) for homogeneous ground conditions, we treated the Fourier coefficients as frequency series, in the other method (Method II) for weekly inhomogeneous ground conditions, we discussed power spectra which were divided into the site effect and random component. Under those method, the stochastic properties of spectral uncertainties were derived theoretically. Furthermore, from this we can evaluate the estimation errors of the stochastically interpolated spectra for Method II. The numerical comparison of the two methods suggests that we can obtain reasonable results through Method II, even though there are some inhomogeneous ground structures.

ACKNOWLEDGMENT: This study was supported in part by the Grants-in-Aid (No.2237 and No.08452235). Being in progress of this study, H. Morikawa was supported by a Fellowships of the Japan Society for the Promotion of Science for Japanese Junior Scientists.

REFERENCES


