Modeling of Indian strong motion data using Empirical Mode Decomposition technique

S Sangeetha
Research Scholar (PhD): Dept. of civil Engineering
Indian Institute of Technology, Madras
Chennai, India
ssangee.civil@gmail.com

STG Raghukanth
Associate Professor: Dept. of civil Engineering
Indian Institute of Technology, Madras
Chennai, India
raghukanth@iitm.ac.in

Abstract—In India, the frequency of occurrence of earthquakes is more, especially in Himalaya, northeast India and Gujarat region. Therefore it becomes essential to estimate the seismic inputs for such earthquakes to reduce the structural damage. From engineering point of view, the most sought-after data is the strong motion accelerograms (SMA), recorded in the places where earthquake has occurred. The article analyzes the Indian strong motion records of past earthquakes by empirical mode decomposition (EMD) technique and in-turn presents few strong ground motion parameters, which finds its use in the simulation of artificial ground motions. The recorded earthquake acceleration time histories are decomposed into a finite number of empirical modes of oscillation. The instantaneous frequency and amplitude of these modes and evolutionary power spectral density (PSD) is estimated from the Hilbert Huang transform (HHT). Strong motion parameters such as spectral and temporal centroid, spectral and temporal standard deviation, instantaneous power, Arias intensity, correlation coefficient of frequency and time are derived from the evolutionary PSD. The variation of these parameters with magnitude and distance of the recording station has been examined. Empirical equations to estimate these ground motion parameters are derived from the strong motion data by regression analysis. These equations can be used by engineers to estimate the design ground motion.

Keywords- Evolutionary Power Spectral Density; Empirical Mode Decomposition; Hilbert Huang Transform; Ground motion prediction equations; design ground motion

I. INTRODUCTION

The earthquakes constitute a rapid onset and the most feared natural hazard, that pose a serious threat of catastrophic disaster across the world. The Indian subcontinent is one of the most seismic prone areas in the world. Some of the India’s most devastating earthquakes in the past which caused heavy causalities and economic damages are the 1897 Assam, 1905 Kangra, 1934 Bihar, 1950 Assam, 1967 Koyina, 1993 Kilarri, 1999 Jabalpur, 2001 Bhuj, 2004 Indian Ocean. In the recent times, the frequency of occurrence of earthquakes in the Himalayas and in the north-east India is relatively high compared to peninsular India (Figure 1). This could be due to the several reasons like the subduction of Indian lithosphere under the Eurasian plate, landslides, soil liquefaction etc and hence the hazard in these parts of India is considered to be severe.

Estimating future hazards of a region constitutes an important problem in earthquake engineering. One way to mitigate the destructive impact of earthquakes is by designing the buildings and infrastructure to be safe in case of a future earthquake, which makes it essential to estimate future ground motion. Also, the success of earthquake resistant design of structures strongly depends on the accuracy with which the future ground motion can be determined at a particular site. Particularly, in the design of major structures and facilities such as important buildings, dams, bridges and nuclear power plants, it is highly desirable to know the ground motion at a specific site that would result from a particular earthquake event. Unfortunately, records of the Indian strong motion data are very limited for engineers to rely upon. Estimation of ground motion time histories under such circumstances by understanding the needs of engineers is a challenging task.

The starting point in the estimation of future ground motion is the modeling of the available strong motion data. Recorded ground motion time histories during the past earthquakes in a region provide valuable information about the expected characteristics of ground motion at that particular site during a future earthquake in that region. These estimates are essential for evaluating earthquake resistant design procedures, estimation of attenuation characteristics, assessment of seismic hazard and earthquake risk. Accordingly, analysis of strong ground motion has been carried out to understand the potential effect of strong shaking during earthquakes using sophisticated mechanistic models [1] [2] and ground motions have also been simulated for the past earthquakes [3][4]. These models would be an obvious choice to estimate ground motion in seismic hazard analysis, but uncertainties exist in choosing inputs regarding fault rupture process for future earthquakes. Also, this approach seems to be effective for the case of scenario earthquake, but it is inapplicable for probabilistic seismic hazard assessment which is based on consideration of multiple earthquakes occurring at different distances from the site. In addition, the occurrence of earthquake is random, hence
difficulties exist in selecting the design ground motion. This issue can be addressed through probabilistic seismic hazard analysis wherein ground motion from all possible magnitude and location combinations are considered in estimating hazard. In such situation, empirical models with few input parameters based on strong motion records would be an obvious choice. These empirical methods are more popular among engineers as it requires very few input parameters in the generation of artificial time histories.

Several parameters required for engineering purposes has been identified from the earthquake acceleration time histories. These parameters also known as strong motion parameters characterize the ground motion. The most popular measures are peak ground acceleration (PGA), peak ground velocity (PGV) and Spectral acceleration (Sa) respectively. Ground motion prediction equations (GMPE) have been developed to obtain these measures based on the recorded strong motion data. These empirical equations are routinely used in linear structural analysis. However, nonlinear dynamic analysis demands acceleration time histories valid in all frequency ranges. For this purpose, stochastic process models for generating samples of acceleration time histories have also been developed [5]. However spectral nonstationarity which has a substantial effect on the structural response[6][7] is ignored in these approaches. Another drawback with these approaches are, most of the above mentioned models are Fourier based. But, it is well known that for nonstationary processes like the earthquake acceleration time histories, whose duration is short and amplitude and frequency contents change with time, Fourier based approaches can yield distorted information.

This has been pointed out by [8] and has shown that adaptive basis functions are required to understand nonlinear and nonstationary data. This technique has been used to analyze the NGA (Next Generation of Ground-Motion Attenuation Models) database (Raghukanth and Sangeetha 2012) and strong motion parameters required for the simulation of acceleration time histories have been arrived. However the developed ground motion prediction equations using the NGA database may not be valid for the Indian scenario. Hence, in this paper the Empirical Mode Decomposition (EMD) technique combined with Hilbert-Huang Transform (HHT) is used to model the Indian strong motion data of past events. Important parameters of the evolutionary PSD are estimated from Hilbert spectral analysis. With these parameters one can generate ensemble number of ground motions, which is the key input in any nonlinear structural analysis.

II. GROUND MOTION DATA

The prerequisite to develop any model is the availability of a good quality strong motion dataset. For the present study the Indian strong motion data is considered. The strong motion data for the past earthquakes (since 2005 onwards) are available in the Pesmos website (http://www.pesmos.in/) as acceleration time histories. Figure 2a shows the fault map of India with topography. The location of strong motion stations is also shown in Figure 2a. This database consists of 430 strong motion records generated by 135 events. There are total of 94 recording stations. Figure 2b shows the earthquake epicenters of all the events in the database. It can be observed from the figure that the moment magnitude (\(M_p\)) of the events lie in between \(M_p\) 2.3 to 7.8. A closer view of the strong motion stations in the north east India along with the faults are available in Figure 2c. The Indian strong motion database is shown in Figure 3(a,b) as a function of magnitude and epicentral distance and hypocentral distance respectively. The epicentral distance (\(R_{epic}\)) varies from 2 km to 1000 km, the hypocentral distance varies from 9 km to 1000 km, where as focal depth of the events lie in between 2 to 190 km. The local soil condition of the recording stations characterized in terms of average shear wave velocity in the top 30 meters (\(V_{s30}\)). This information at all the strong motion stations is available in the PESMOS website (http://www.pesmos.in/) along with data. Since horizontal components are widely used in seismic design, both the north-south and east-west components are used in this study to estimate ground motion parameters.

![Figure 1. Historic Seismicity superposed on known Faults (38860 events of Mw ≥4 including foreshocks and aftershocks)](image)

III. EMPIRICAL MODE DECOMPOSITION

The adaptive technique proposed by [8] known as Hilbert Huang transform (HHT) consists of empirical mode decomposition (EMD) and Hilbert transform. The EMD decomposes the signal into a finite and often small number of empirical modes called intrinsic mode functions (IMFs). EMD-HHT technique has found a lot of applications especially in earthquake engineering. [9] presents a detailed review of the EMD algorithm to extract the IMFs. The recent Sikkim earthquake,2011 caused great damages to infrastructure. The earthquake was recorded at 13 station. However, Gangtok station which is 80km from the epicenter of the sikkim earthquake recorded the highest PGA of nearly 200cm/sec². Hence in the following sections emphasis is made on the Gangtok station, since it is becomes necessary to design the structures for the maximum PGA that can be obtained during a future earthquake at that site. In Figure 4(a,b), IMFs of both horizontal components acceleration time histories recorded at Gangtok station during the Sikkim earthquake (\(M_p\) 6.8) are shown. The sum of the IMFs leads to the original data. After
the EMD, the earthquake acceleration time history can be expressed in terms of IMFs as follows, \( a(t) = \sum \text{IMF}_i(t) \)

Figure 2a. Location of the 94 strong motion station used for the present study

Figure 2b. Epicenters of the 430 events in the database

Figure 2c. Fault Map of NorthEast India along with the stations
IV. HILBERT HUANG TRANSFORM (HHT)

The evolutionary spectrum can be constructed by taking the Hilbert transform the IMFs as

\[ D_j(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \text{IMF}_j(\tau) \frac{d\tau}{\tau - t} \]

where \( \text{IMF}_j \) and \( D_j \) form the complex conjugate pair, one can form the analytical function as

\[ Z_j(t) = \text{IMF}_j(t) + iD_j(t) = C_j(t)e^{\omega_j t} \]

The instantaneous amplitude, \( C_j(t) \) of the \( j \)th IMF is defined as

\[ C_j(t) = \sqrt{\text{IMF}_j(t)^2 + D_j(t)^2} \]

and the instantaneous phase can be expressed as

\[ \theta_j(t) = \tan^{-1}\left( \frac{D_j(t)}{\text{IMF}_j(t)} \right) \]

The instantaneous frequency (IF) of the \( j \)th IMF can be estimated as

\[ \omega_j(t) = \frac{d\theta_j(t)}{dt} \]

The original acceleration time history can be expressed as the real part of the sum of the Hilbert transform of each IMF

\[ a(t) = \text{Re}\left\{ \sum_{j=1}^{n} C_j(t)e^{i\theta_j(t)} \right\} \]

The estimated instantaneous frequencies of the recorded acceleration time histories recorded at Gangtok station for the Sikkim event are shown in figure 5(a,b). It can be observed from the figure that the instantaneous frequencies of the IMFs decrease in the higher modes. The mean frequency of each IMF is determined by averaging instantaneous frequency over time. In a similar fashion, the central frequency of the 10 IMFs for all the 430 acceleration time histories have been estimated by HHT. Earthquake ground motion in general can be decomposed into number of IMFs. As the variance of the higher modes is insignificant, only the first 10 IMFs are considered for further statistical analysis. Table 1 shows the mean and standard deviation of the mean frequency estimated from 430 time histories for both the horizontal components. It is observed that the mean of the instantaneous frequency remains constant and doesn’t show any pattern with \( M_w \), \( R \).

The percentage of variance explained by each IMF, or the contribution of each IMF to total variance of the data is determined for all the 430 time histories. The mean and standard deviation of the percentage of variance estimated from 430 samples is presented in Table 1 for both EW and NS components. It is observed that IMF$_1$ with a central frequency of 18.64 Hz is the predominant mode contributing to 43 % of the variability in the data. IMF$_2$ with a central frequency of 8.39 Hz is the second most important mode in both the cases. The next important modes are IMF$_3$ and IMF$_4$ oscillating about 4
Hz and 2 Hz contributing to 5% variability in the data. These four dominant modes explain about 97% variability in both the (East-West) EW and (North-South) NS component of the acceleration time histories.

Figure 4a. Intrinsic Mode Functions of acceleration recorded at Gangtok station during Sikkim earthquake

The percentage of variance explained by each IMF, or the contribution of each IMF to total variance of the data is determined for all the 430 time histories. The mean and standard deviation of the percentage of variance estimated from 430 samples is presented in Table 1 for both EW and NS components. It is observed that IMF\(_1\) with a central frequency of 18.64 Hz is the predominant mode contributing to 43% of the variability in the data. IMF\(_2\) with a central frequency of 8.39 Hz is the second most important mode in both the cases. The next important modes are IMF\(_3\) and IMF\(_4\) oscillating about 4 Hz and 2 Hz contributing to 5% variability in the data. These four dominant modes explain about 97% variability in both the EW and NS component of the acceleration time histories.

V. EVOLUTIONARY POWER SPECTRAL DENSITY

The evolutionary PSD \([G(t, \omega)]\) of the acceleration time history can be constructed from HHT as [10]

\[
G(t, \omega) = \sum_{j=1}^{n} \frac{1}{2} \delta(\omega - \omega_j) \left| C_j^2(t) \right|
\]  

(7)

The corresponding evolutionary PSD of both EW and NS component is shown in figures 7(a,b). The instantaneous average power \([Pa(t)]\), the central frequency \([Fc(t)]\) and the frequency bandwidth \([Fb(t)]\) can be estimated from the first three spectral moments as [11]

\[
Pa(t) = \int_0^{\infty} G(t, \omega) \ d\omega
\]  

(8)

\[
F_c(t) = \frac{\int_0^{\infty} \omega G(t, \omega) \ d\omega}{\int_0^{\infty} G(t, \omega) \ d\omega}
\]  

(9)

\[
F_b(t) = \sqrt{\frac{\int_0^{\infty} \omega^2 G(t, \omega) \ d\omega - \left( \int_0^{\infty} \omega G(t, \omega) \ d\omega \right)^2}{\int_0^{\infty} G(t, \omega) \ d\omega}}
\]  

(10)

These three strong motion parameters are shown in figure 8(a,b) for both the horizontal components recorded at Gangtok station for the 2011 Sikkim earthquake. It can be observed that the central frequency and frequency
bandwidth oscillate around a constant mean and doesn’t show any pattern with time where as instantaneous power decreases with time.

Figure 5. Instantaneous frequencies of acceleration time histories shown in figure 4(a,b)

TABLE 1 CENTRAL FREQUENCY OF THE IMF’S IN Hz AND % VARIANCE CONTRIBUTED TO TOTAL VARIABILITY OF THE DATA

<table>
<thead>
<tr>
<th>IMF</th>
<th>Component: East</th>
<th>Component: North</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency Mean±STD</td>
<td>%Variance Explained Mean±STD</td>
</tr>
<tr>
<td>IMF1</td>
<td>18.64±6.71</td>
<td>42.56±21.64</td>
</tr>
<tr>
<td>IMF2</td>
<td>8.39±3.49</td>
<td>34.76±13.35</td>
</tr>
<tr>
<td>IMF3</td>
<td>4.23±1.71</td>
<td>15.62±10.77</td>
</tr>
<tr>
<td>IMF4</td>
<td>2.21±0.95</td>
<td>4.67±4.99</td>
</tr>
<tr>
<td>IMF5</td>
<td>1.14±0.46</td>
<td>1.13±1.75</td>
</tr>
<tr>
<td>IMF6</td>
<td>0.58±0.25</td>
<td>0.34±0.99</td>
</tr>
<tr>
<td>IMF7</td>
<td>0.28±0.11</td>
<td>0.12±0.13</td>
</tr>
<tr>
<td>IMF8</td>
<td>0.13±0.06</td>
<td>0.09±0.24</td>
</tr>
<tr>
<td>IMF9</td>
<td>0.06±0.02</td>
<td>0.08±0.52</td>
</tr>
<tr>
<td>IMF10</td>
<td>0.03±0.01</td>
<td>0.04±0.16</td>
</tr>
</tbody>
</table>

Figure 6a. Central frequency of all the IMF's as a function of Magnitude, rupture distance (Component: EW)
Figure 6b. Central frequency of all the IMF's as a function of Magnitude, rupture distance (Component: NS)

Figure 7(a,b). Evolutionary PSD of acceleration time histories shown in figure 4(a,b) for the horizontal component

Figure 8a. Instantaneous power, Central frequency and Frequency bandwidth estimated from evolutionary PSD . (Component: EW)
VI STRONG MOTION PARAMETERS

A few parameters which are of engineering interest can be identified from the evolutionary PSD. In this study, six parameters namely total energy of the acceleration time history ($E_{acc}$) also known as Arias Intensity ($AI$), spectral centroid [$E(f)$], temporal centroid [$E(t)$], spectral standard deviation [$S(f)$], temporal standard deviation [$S(t)$] and the correlation of time and frequency [$\rho(t,f)$] are used to characterize the evolutionary PSD of all the 430 acceleration time histories. These parameters are estimated from the evolutionary PSD as

$$E_{acc} = \int_{0}^{\infty} \int_{0}^{\infty} G(t,\omega) \, d\omega \, dt$$  \hspace{0.5cm} (11)

$$E(\omega) = \frac{\int_{0}^{\infty} \omega G(t,\omega) \, d\omega \, dt}{\int_{0}^{\infty} G(t,\omega) \, d\omega \, dt}$$  \hspace{0.5cm} (12)

$$S^2(\omega) = \frac{\int_{0}^{\infty} (\omega - E(\omega))^2 G(t,\omega) \, d\omega \, dt}{\int_{0}^{\infty} G(t,\omega) \, d\omega \, dt}$$  \hspace{0.5cm} (13)

$$E(t) = \frac{\int_{0}^{\infty} \int_{0}^{\infty} G(t,\omega) \, d\omega \, dt}{\int_{0}^{\infty} G(t,\omega) \, d\omega \, dt}$$  \hspace{0.5cm} (14)

$$S^2(t) = \frac{\int_{0}^{\infty} (t - E(t))^2 G(t,\omega) \, d\omega \, dt}{\int_{0}^{\infty} G(t,\omega) \, d\omega \, dt}$$  \hspace{0.5cm} (15)

$$\rho(t,\omega) = \frac{\int_{0}^{\infty} \int_{0}^{\infty} (t - E(t))(\omega - E(\omega)) G(t,\omega) \, d\omega \, dt}{\int_{0}^{\infty} \int_{0}^{\infty} (t - E(t))^2 G(t,\omega) \, d\omega \, dt \int_{0}^{\infty} \int_{0}^{\infty} (\omega - E(\omega))^2 G(t,\omega) \, d\omega \, dt}$$  \hspace{0.5cm} (16)

These parameters are estimated for both the horizontal components. Figure 9(a,b) shows the spectral centroid and its standard deviation as a function of magnitude and distance for both the horizontal components. It can be observed that spectral centroid decreases with increase in magnitude and closest distance to rupture. The temporal parameters are shown in figure 10(a,b) for both the horizontal components. The temporal centroid and standard deviation increases with increase in magnitude and distance to rupture. Figure 11 shows the variation of Arias intensity and correlation coefficient of time and frequency with the magnitude and distance.
Figure 9(a,b). Spectral centroid and Spectral standard deviation of the evolutionary PSD as a function of Mw, R

Figure 10(a,b). Temporal centroid and Temporal standard deviation of the evolutionary PSD as a function of Mw and R.

Figure 11(a,b). Arias intensity and correlation coefficient of the evolutionary PSD as a function of Mw and R.
VII GROUND MOTION PREDICTION EQUATIONS

Characterization of the evolutionary PSD is followed by the development of empirical equations to predict the six strong motion parameters. These attenuation equations are the key component in seismic hazard analysis. Several functional forms of ground motion attenuation have been proposed in the literature reflecting salient aspects of the spread of ground motion ([12], [13], [14]). After reviewing the various available forms of equations, it has been decided to develop the empirical relation for both horizontal components in the form

\[
\ln(y) = \beta_1 + \beta_2 M_w + \beta_3 \ln(M_w) + \beta_4 \exp(M_w) + \beta_5 (\sqrt{R^2 + h^2}) + \beta_6 \ln(V_{s30}) + \ln(\varepsilon)
\]  

(17)

where y, M and R refer to ground motion parameter, moment magnitude and distance respectively. h, is the pseudo focal depth to control the ground motion parameters close to the fault rupture. \(\ln(\varepsilon)\) is the error associated with the regression. A two stage regression analysis is carried out on the SMA data to obtain the coefficients. In this procedure equation 17 is first written as

\[
\ln(y) = \sum_{i=1}^{n} E_i + \beta_2 (\sqrt{R^2 + h^2}) + \beta_3 \ln(V_{s30}) + \varepsilon
\]  

(18)

Here \(n\) is the number of earthquakes = 430, \(E_i\) is a constant for the \(i^{th}\) earthquake and \(li\) is a dummy variable which takes value 1 for the \(i^{th}\) earthquake and 0 otherwise. \(\varepsilon\) is the error in the regression associated with distance dependence. The unknowns (\(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\)) in the above equation are determined from linear regression analysis. The obtained (\(\beta_1, \beta_2, \beta_3, \beta_4\)) are reported in Tables 2 and 3. Once the distance and soil coefficients are known, the coefficients (\(\beta_1, \beta_2, \beta_3, \beta_4\)) are determined from \(E_i\) as,

\[
E_i = \beta_1 + \beta_2 M_w + \beta_3 \ln(M_w) + \beta_4 \exp(M_w) + \beta_5 (\sqrt{R^2 + h^2}) + \beta_6 \ln(V_{s30}) + \varepsilon
\]  

(19)

where \(\varepsilon\) is the error in the regression associated with magnitude dependence. The constants of regression are reported in Tables 2 and 3. The standard deviation of the ground motion which is required in seismic hazard analysis can be expressed as

\[
\sigma(\ln(\varepsilon)) = \sqrt{\sigma(\varepsilon_{mean})^2 + \sigma(\varepsilon)^2}
\]

(20)

These values are also reported in the above tables. These results can be used to construct the mean and (mean+sigma) ground motion parameters of the evolutionary PSD on any site condition in shallow active regions in the world. Figure 12 shows the variation of Arias Intensity with distance for different magnitudes for both the horizontal components.

<table>
<thead>
<tr>
<th>Table 2. Coefficients in equation 17 for NS component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component: NS</td>
</tr>
<tr>
<td>Eacc (cm²/sec²)</td>
</tr>
<tr>
<td>4.0931</td>
</tr>
<tr>
<td>3.2951</td>
</tr>
<tr>
<td>2.6539</td>
</tr>
<tr>
<td>3.9698</td>
</tr>
<tr>
<td>1.0460</td>
</tr>
<tr>
<td>0.0598</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3. Coefficients in equation 17 for EW component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component: EW</td>
</tr>
<tr>
<td>Eacc (cm²/sec²)</td>
</tr>
<tr>
<td>4.2</td>
</tr>
<tr>
<td>3.3</td>
</tr>
<tr>
<td>2.8</td>
</tr>
<tr>
<td>4.1</td>
</tr>
<tr>
<td>1.2</td>
</tr>
<tr>
<td>0.07</td>
</tr>
</tbody>
</table>
VIII SUMMARY AND CONCLUSION

This article explores the application of EMD-HHT technique for characterizing earthquake acceleration time histories. A total of 430 horizontal components generated by 135 events from Indian strong motion database (Pesmos) have been used in the analysis. The EMD technique shows some interesting features of SMA data. It is observed that earthquake accelerations can be represented as a sum of ten independent modes. The mean instantaneous frequency of these modes estimated from time averaging doesn't show any pattern with magnitude, distance to rupture and the site shear wave velocity.

The contribution of the Individual modes to the total variability of the data is reported in Table 1. The first IMF with mean instantaneous frequency of 18 Hz explains the maximum variability of the data in both the horizontal components. The evolutionary PSD of the strong motion data is constructed from HHT. Since modeling the evolutionary PSD is complex, spectral and temporal centroid, spectral and temporal standard deviation, Arias intensity and correlation coefficient have been extracted from the PSD. The variation of these parameters with magnitude and distance has been reported in figures 9-11. The spectral centroid decreases with increase in magnitude and rupture distance. The temporal centroid which characterizes duration of the ground motion increases with magnitude and rupture distance. These patterns are consistent with the seismological concepts. Empirical equations to predict these six parameters have been derived by two-step regression analysis for both the horizontal components. These equations can be used to estimate the mean values of the six ground motion parameters for a given magnitude, rupture distance and the site condition. The six strong motion parameters estimated in this study can be used as an input in these studies and an ensemble of acceleration time histories can be constructed by the spectral representation method of [10].

REFERENCES