

SEISMIC INPUT FOR CHENNAI USING ADAPTIVE KERNEL DENSITY ESTIMATION TECHNIQUE

G. R. Dodagoudar

Associate Professor, Indian Institute of Technology Madras, Chennai - 600036, goudar@iitm.ac.in

P. Raguathan

Formerly Post Graduate Student, Indian Institute of Technology Madras, Chennai - 600036,
parimala.raguathan@gmail.com

C. K. Ramanna

Formerly Research Scholar, Indian Institute of Technology Madras, Chennai - 600036,
ce07d001@smail.iitm.ac.in

Abstract

Seismic input at a particular site can be estimated quantitatively using probabilistic or deterministic approach. Probabilistic seismic hazard analysis (PSHA) provides a framework in which uncertainties in the size, location, rate of recurrence and effects of earthquakes are explicitly considered in the evaluation of seismic hazard. The probabilistic way of analyzing the seismic hazard was developed conventionally by introducing zones in the seismogenic regions based on regional seismotectonic and geologic setting. The seismic uniformity is assumed within these source zones. Later, many researchers found that the conventional approach has many drawbacks viz., difficulty in delineating seismic sources into various zones, difficulty in applying Gutenberg-Richter (G-R) recurrence relationship to characterize the seismic source for low seismicity regions and distributed seismicity, and the consideration of uniform seismicity within the zone is also questionable. Because of these issues, several alternative methods to hazard estimation have been proposed in the literature. In the present study, zone free approach is proposed to evaluate the spatial distribution of seismicity based on kernel density estimation technique. The kernel technique provides a spatial variation of the seismic activity rate unlike the conventional approach where it is constant for a seismic source zone. The fixed bandwidth kernel poorly evaluates the earthquake distributions since the earthquake catalogue has several areas of high activity clusters and low background seismicity. Therefore in this study, *clustering based adaptive kernel technique* is proposed to find the spatial activity rate and integrated with other forms of uncertainty in magnitude and distance to determine the probability of exceedance of the selected ground motion parameter. The proposed methodology of seismic hazard analysis has been used for Chennai, southern India and the seismic input is provided in the form of Peak Ground Acceleration (PGA) and Uniform Hazard Spectra (UHS) for return periods of 475 and 975 years. The UHS obtained are compared with the Cornell-McGuire approach and IS 1893: 2002.

Keywords: Probabilistic seismic hazard, Adaptive kernel technique, Clustering method, PGA, UHS.

1. INTRODUCTION

Probabilistic seismic hazard analysis (PSHA) provides a framework in which uncertainties in the size, location, rate of recurrence and effects of earthquakes are explicitly considered in the evaluation of seismic hazard. The probabilistic way of analyzing the seismic hazard was developed conventionally by introducing zones in the seismogenic regions based on regional seismotectonic and geologic setting for distributed or diffused seismicity region. The conventional way of characterizing the seismicity for such regions has the following drawbacks. It needs the knowledge of expertise to delineate the seismic source into zones. The assumption of homogenous seismicity in each source zone is questionable and the applicability of G-R recurrence law for low to moderate seismicity region is also questionable (Beauval et al., 2006). Bender (1986) has explained that there would be an abrupt change in the seismicity at the zonal boundaries. Hence to overcome all these disadvantages, in the present study, the zone free approach (i.e., Regionalization free approach) is proposed to evaluate the spatial distribution of seismicity (which replaces the seismic activity rate of G-R recurrence law) based on kernel density estimation (KDE) technique. In specific, clustering based adaptive kernel density estimation (AKDE) technique has been adopted since it has the advantage in case of multimodal distributions and smoothing the long tail distributions as compared to fixed kernel density estimation (FKDE) technique (Ramanna and Dodagoudar, 2011). The proposed methodology is applied to Chennai and the results are provided in the form of peak ground acceleration (PGA) and uniform hazard spectra (UHS). The application of zone free method is justified for Chennai for the reason that it falls under distributed seismicity region where the geological features causing earthquakes are difficult to determine. This is especially true for southern part of Peninsular India (PI) from 20°N latitude and down.

2. ADAPTIVE KERNEL DENSITY ESTIMATION TECHNIQUE

The kernel density estimation consists of determining the probability density function (PDF) $f(x)$ by placing standard form of distributions such as Uniform, Triangular, Normal or Epanechnikov density curves on the sample or data points known as kernels. The function $f(x)$ is then determined as the normalized sum total of these kernels (Figure 1). A multivariate PDF determined by the KDE is of the form:

$$f(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i^d} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h_i}\right) \quad (1)$$

where n is the number of sample data \mathbf{x}_i , h_i is the variable bandwidth equal to $b_i H$, where b_i is the local bandwidth factor and H is the global bandwidth, \mathbf{x} is the estimation or evaluation point where density is determined and $K(\cdot)$ is the kernel of any form. The multivariate normal kernel is

$$K(t) = (2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2} t^T t} \quad (2)$$

in which $t = (\mathbf{x} - \mathbf{x}_i)/h_i$, T stands for transpose and d is d -dimensional space.

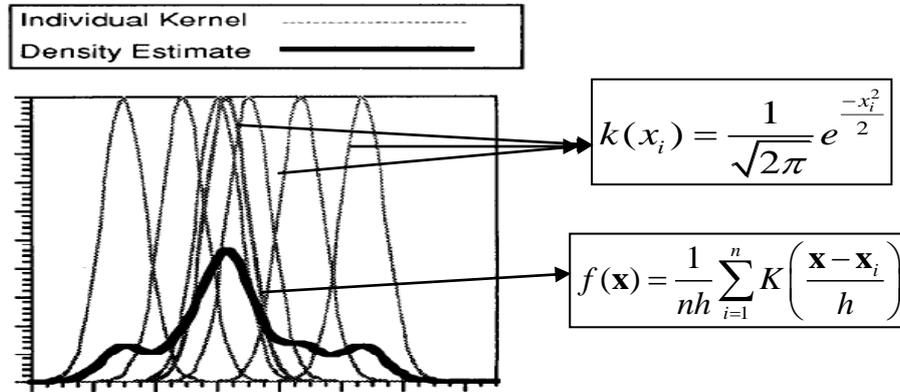


Figure 1: Graphical representation of univariate KDE technique

2.1 Adaptive Bandwidth by Clustering Technique and Nearest Neighborhood Method

In this study, the clustering technique is adopted to determine the local bandwidth factor where hierarchical clustering procedure is used. The approach of using clustering technique for the determination of local bandwidth factor was proposed by Wu et al. (2007). Two clusters are taken at a time and merged using average linkage method, where the average distances between pairs of members in the respective sets are found. The algorithm to find the local bandwidth factors is available in the literature of Johnson and Wichern (2007). The local bandwidth factor is given as

$$b_i = \frac{1}{n} \sum_{j=1}^{n_i} l_j \quad (3)$$

where n is the total number of times that a cluster containing x_i is merged into a large cluster (i.e., total number of merges that involve x_i) and l_1, l_2, \dots, l_n are the distance levels at which n merges take place.

There are several methods to determine the global bandwidth (e.g., Silverman, 1986). For hazard analysis, the bandwidth for every magnitude bin is determined using nearest neighborhood method by force fitting power law (Woo, 1996) and is given as

$$H(M) = ce^{(dM)} \quad (4)$$

where c and d are the bandwidth parameters. These parameters are calculated by forming various magnitude bins and for each earthquake event within the bin, the distance to the nearest epicentre is determined. The mean nearest distance for each bin is obtained and through a least-square fit between the magnitude and mean nearest distance, the parameters are evaluated.

3. AKDE TECHNIQUE FOR SEISMIC HAZARD ANALYSIS

The PSHA using AKDE technique has two parts. Firstly determine the seismic activity rate density function $\nu(M, \mathbf{x})$ using adaptive kernel technique which uses clustering procedure and nearest neighborhood method for determining adaptive bandwidth. In the kernel technique to PSHA, the seismic activity rate ν_i (conventionally defined in terms of G-R recurrence law) is replaced by a spatially varying activity rate which is given as

$$\nu(M, \mathbf{x}) = \sum_{i=1}^n \frac{K(M, \mathbf{x} - \mathbf{x}_i)}{T_i} \quad (5)$$

where n is the number of earthquake events, $(\mathbf{x} - \mathbf{x}_i)$ is the distance to the epicentre and T_i is the effective return period, given as

$$T_i = \sum_i p_i D_i \quad (6)$$

where p_i is the user assigned detection probability of an earthquake event based on the seismicity of the region in a particular time period D_i .

The $K(M, \mathbf{x} - \mathbf{x}_i)$ is determined using Vere-Jones (1992) anisotropic multivariate kernel and is of the form:

$$K(M, r) = \frac{n-1}{\pi h^2(M)} \frac{1 + \delta \cos^2 \phi}{1 + (\delta/2)} \left(1 + \left(\frac{r}{h(M)} \right)^2 \right)^{-n} \quad (7)$$

where n is the exponent of the power law or also known as fractal scaling index, the value of n lying between 1.5 and 2 and has little effect on the hazard results (Molina et al., 2001), r is the distance to the epicentre $(\mathbf{x} - \mathbf{x}_i)$, the parameter ϕ is the angle subtended at r between the intersection of the fault plane with the Earth's surface and the epicenter location and δ is the degree of anisotropy, having value 0 and above. A value of zero indicates isotropy and higher value signifies anisotropy. In the adaptive kernel technique, h is a function of both the magnitude and space i.e., $h(M, \mathbf{x}_i)$ hence $K(\cdot)$ varies spatially for each magnitude M .

In the second part of PSHA, the annual rate of exceedance is calculated by clubbing the spatial activity rate with uncertainties in magnitude, location and ground motion. The mean annual rate of exceedance λ_{y^*} of the selected ground motion parameter Y exceeding a particular value y^* is given as

$$\lambda_{y^*} = \sum_{i=1}^{N_S} \sum_{j=1}^{N_M} \sum_{k=1}^{N_R} \nu(M, \mathbf{x}) P[Y > y^* | m_j, r_k] P[M = m_j] P[R = r_k] \quad (8)$$

where N_S is the number of sources, N_M is the range of magnitudes, N_R is all the possible range of distances from site to source, $P[Y > y^* | m_j, r_k]$ is obtained from the attenuation relationship, $P[M = m_j]$ and $P[R = r_k]$ are obtained from the probability density function of magnitude and distance respectively. The uncertainty in distance is accounted for by smearing the activity rate over a specified location error for epicenter. For uncertainty in magnitude, a normal distribution is assumed. The uncertainty in attenuation relationship is treated in the same manner as in the conventional Cornell-McGuire approach.

4. HAZARD ANALYSIS FOR CHENNAI - RESULTS

Chennai city (13.0833°N, 80.2833°E) lies in the southern Peninsular India (PI) which is a Precambrian stable continental region (SCR). This part of PI is known for its distributed seismicity and the earthquakes caused are due to intraplate stress within the pre-existing weak zones. The geological and seismotectonic setting around Chennai for an influence area of 300 km radius is shown in Figure 2.

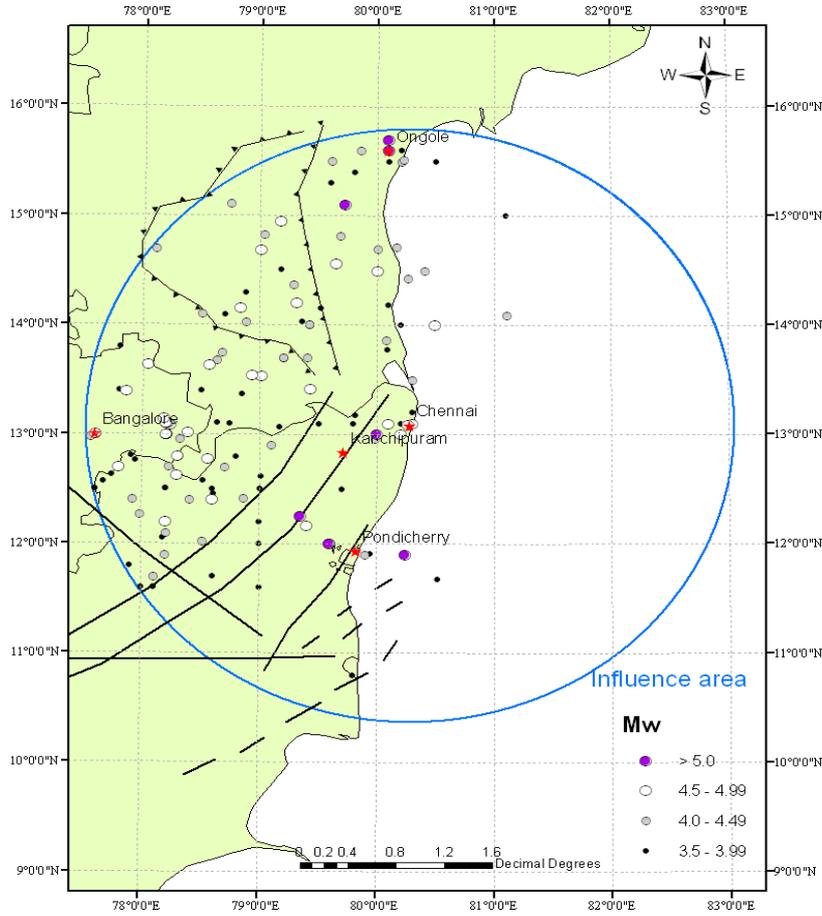


Figure 2: Geological and seismotectonic setting around Chennai

The earthquake data was compiled from various sources and well documented in Rangunathan (2011) for Chennai region. A total of 173 earthquakes of $M_w \geq 3.5$ were compiled for a circular influence area of 300 km radius around Chennai from the year 1507 to 2009 A.D. Gardner and Knopoff (1974) dynamic windowing technique is used to remove the fore and after shocks which resulted in 151 Poissonian events. The attenuation relationship suggested by Iyengar et al. (2010) for South India is used in the hazard analysis. The functional form of this attenuation relationship is

$$\ln(S_a / g) = c_1 + c_2 M + c_3 M^2 + c_4 r + c_5 \ln(r + c_6 e^{c_7 M}) + c_8 \log(r) f_0 + \ln(\varepsilon) \quad (9)$$

A single seismic area source zone with an average hypocentral depth of 17 km was considered for the analysis. The total study area of 600 km \times 600 km i.e., 300 km – control region around Chennai (13.08° N, 80.28° E) is divided into 10 km \times 10 km grids. The kernel method of hazard analysis will not account for the occurrence of magnitudes greater than the historical maximum magnitude unless the uncertainties in magnitude determinations are added. The results of the kernel method greatly depend on the value of the uncertainty

considered for binning the earthquake catalogue. The magnitude bins are formed by considering uncertainty in magnitude as ± 0.49 according to Woo (1996). The value of local bandwidth factor b is determined using clustering method. Figure 3 shows the distribution of epicenters around Chennai for the magnitude bin 3.51-4.49. The diameter of each circle indicates qualitatively the value of local bandwidth factor at the corresponding epicenter. It can be observed from the figure that, in the region of highly clustered epicenters, smaller local bandwidth factor is obtained due to clustering so that the corresponding density will be high and vice versa. The global bandwidth H is estimated using nearest neighborhood method and the values of the bandwidth parameters c and d for Chennai region are found to be 1.266 and 0.623 (Figure 4). The mean nearest distances for all the magnitude bins are given in Table 1. In the kernel methodology to PSHA, the spatial activity rate $v(M, \mathbf{x})$ is the function of both the magnitude and space, where the local bandwidth factor controls the spatial smoothing process as a spatial variant and the global bandwidth controls the spatial smoothing process as a magnitude variant as in Eq. (4).

Distribution of earthquake epicenters for M_w 3.51 - 4.49

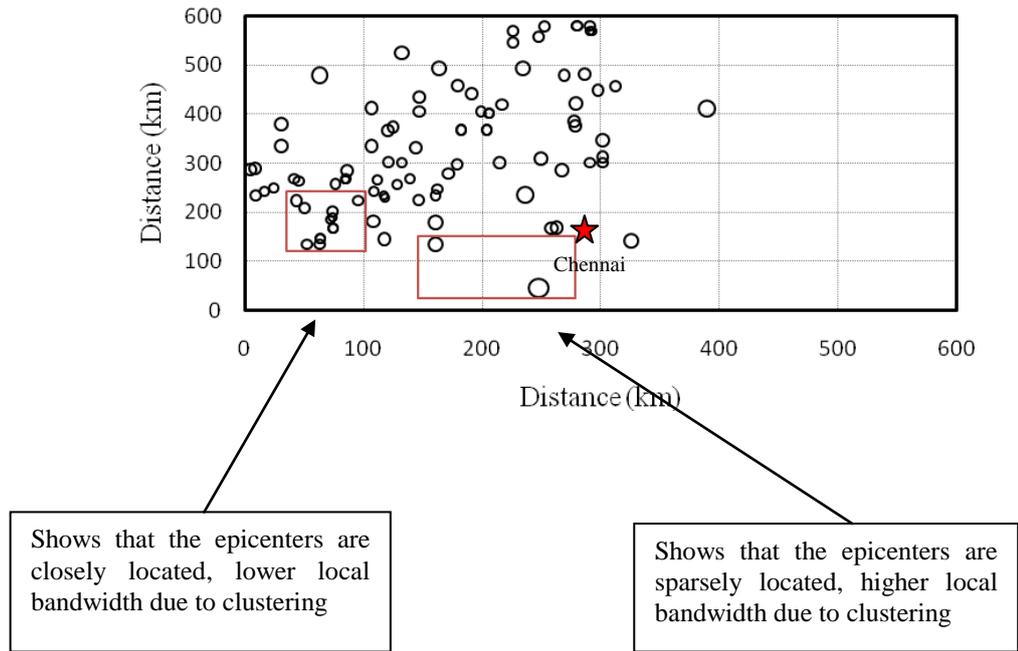


Figure 3: Spatial distribution of epicenters for bandwidth determination

Figure 4: Bandwidth parameters

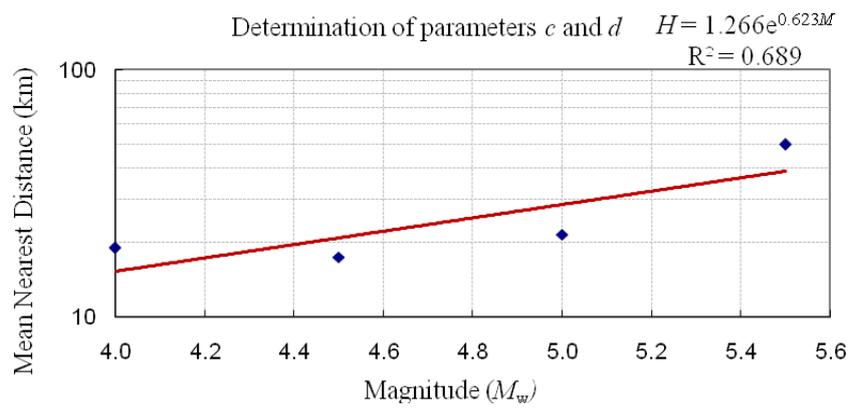


Table 1: Values of mean nearest distance for each magnitude bin

<i>Magnitude bin</i>	<i>Magnitude</i>	<i>Mean nearest distance (km)</i>
4.0 ± 0.49	4.0	18.980
4.5 ± 0.49	4.5	17.426
5.0 ± 0.49	5.0	21.598
5.5 ± 0.49	5.5	49.920

The spatial variation of the activity rate [Eq. (5)] was obtained using Eq. (7) and the effective return period from minimum magnitude $M_w = 4.0$ to maximum catalogue magnitude of $M_w = 5.5$. The reference year (= current year – effective return period) is determined for various magnitude ranges for both the onshore and offshore earthquakes (Table 2). Figure 5 shows the spatial variation of the activity rate for magnitude bin 4.0 which resulted in multimodal distribution. Figure 6 shows the spatial variation of seismic activity rate for magnitude bin 5.0 which resulted in unimodal distribution.

Table 2: Reference year

<i>Magnitude range</i>	<i>Reference year</i>	
	<i>Onshore</i>	<i>Offshore</i>
>5.5	1862	1896
$5.0 - 5.49$	1874	1897
$4.5 - 4.99$	1885	1909
$4.0 - 4.49$	1897	1926
$3.5 - 3.99$	1909	1954

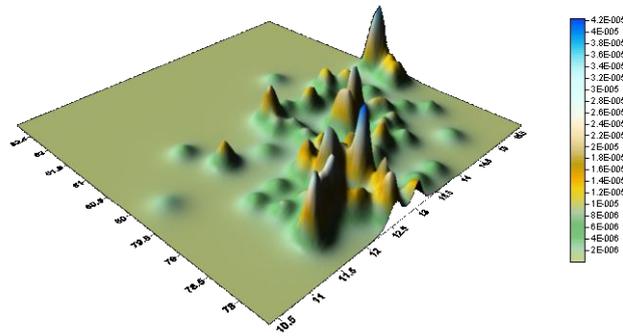


Figure 5: Spatial activity rate for magnitude bin 4.0 (3.51 – 4.49) from adaptive kernel technique

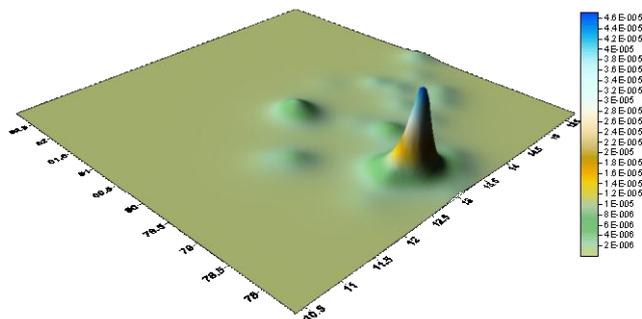


Figure 6: Spatial activity rate for magnitude bin 5.0 (4.51 – 5.49) from adaptive kernel technique

The probability of exceedance was then determined [Eq. (8)] by combining all other forms of uncertainty. The annual probability of exceedance curve and UHS obtained are shown in Figures 7 and 8 respectively.

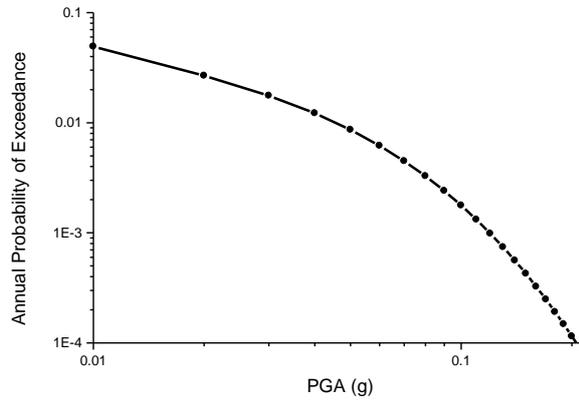


Figure 7: Annual probability of exceedance by clustering based AKDE technique

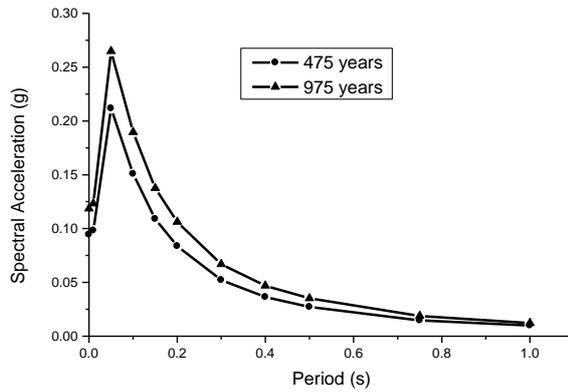


Figure 8: UHS by clustering based AKDE technique

The methodology of PSHA using Cornell-McGuire approach is implemented using CRISIS 2007 (Ordaz et al., 2007). The catalogue completeness analysis was carried out using Stepp's method (1973) considering a controlling area of 300 km radius as one single source zone and the results are given in Table 3. The seismicity parameters obtained from the G-R recurrence law are given in Table 4 and Figure 9. The annual rate of exceedance curve and UHS obtained are shown in Figures 10 and 11 respectively. The comparative plots of UHS obtained by the AKDE technique, Cornell-McGuire approach for 475 years return period and response spectra as given in IS 1893 (Part 1) 2002, design basis earthquake (DBE) for rock site are shown in Figure 12.

Table 3: Completeness parameters

M_w	Completeness period	Completeness year
3.5 - 3.99	40	1968
4.0 - 4.49	40	1968
4.5 - 4.99	50	1958
> 5.0	209	1800

Table 4: G-R recurrence law parameters

Parameter	Value
$\lambda(M_0 = 4)$	2.084
a	5.191
b	1.218
α	11.955
β	2.805

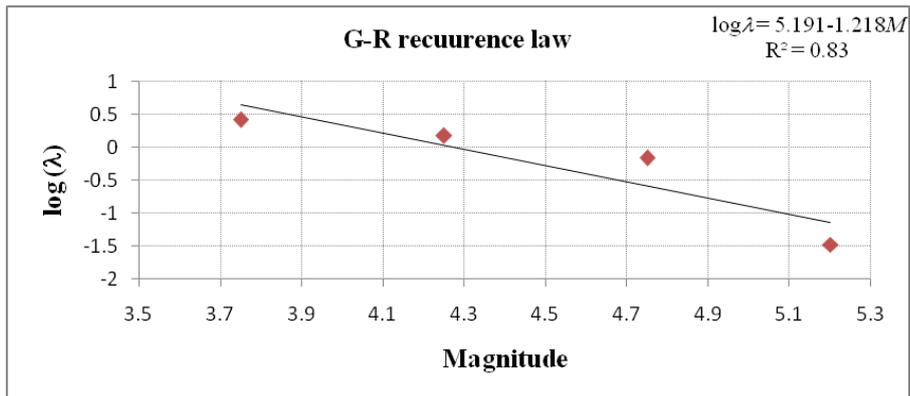


Figure 9: G-R recurrence law for Chennai

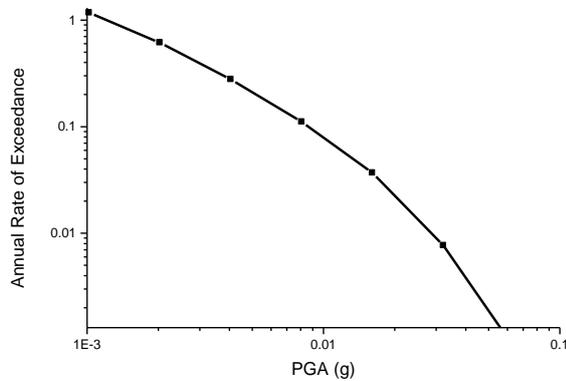


Figure 10: Annual rate of exceedance by Cornell-McGuire approach

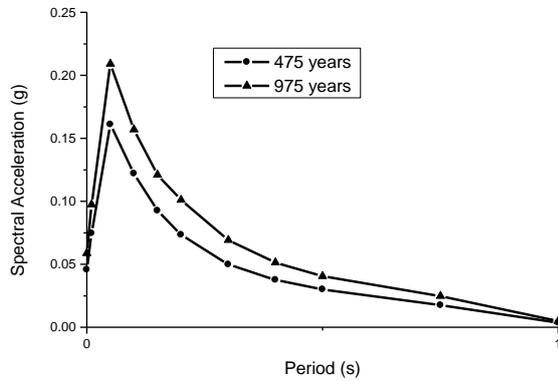


Figure 11: UHS by Cornell-McGuire approach

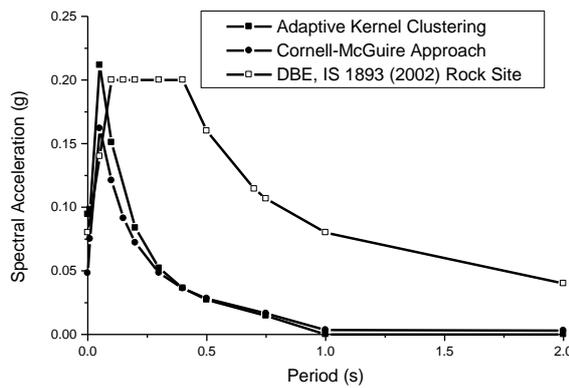


Figure 12: Comparison of UHS for 475 years return period

5. CONCLUSIONS

Seismic hazard analysis is the primary tool to support both the building codes and preparedness. The proposed clustering based adaptive kernel density estimation (AKDE) technique is used to estimate the spatial activity rate density function required in the PSHA. This technique utilizes the clustering procedure so that the denseness and sparseness nature of the earthquake epicenter distributions are clearly captured. The AKDE technique has resulted in the PGA value of 0.094g for Chennai and it is 17.5% more than the IS code specified value (0.08g) for design basis earthquake (DBE). The UHS obtained using the AKDE technique are compared with those of the Cornell-McGuire approach and fixed kernel method for 475 years return period (Figure 12). The AKDE technique yields higher value of hazard when compared to the Cornell-McGuire approach where the difference in spectral acceleration is 0.05g which is approximately 31.5%. It is noted that the tedious job of forming the area source zone which satisfies the homogeneity condition and fitting of the G-R recurrence relationship can be overcome without compensating the accuracy in the case of kernel methods.

References

- Beauval C, Scotti O and Bonilla F (2006) "The Role of Seismicity Models in Probabilistic Seismic Hazard Estimation: Comparison of a Zoning and a Smoothing Approach." *Geophysical Journal International* **165**: 584-595.
- Bender B (1986) "Modelling Source Zone Boundary Uncertainty in Seismic Hazard Analysis." *Bulletin of the Seismological Society of America* **76(2)**: 329-341.
- Gardner J K and Knopoff L (1974) "Is the Sequence of Earthquakes in Southern California, with Aftershocks Removed, Poissonian?." *Bulletin of the Seismological Society of America* **64(5)**: 1363-1367.
- Indian Standards (IS 1893: 2002) (2002) *Indian Standard Criteria for Earthquake Resistant Design of Structures, Part 1 – General Provisions and Buildings*, BIS, New Delhi.
- Iyengar R N, Chadha R K, Balaji Rao K and Raghu Kanth S T G (2010) *Development of Probabilistic Seismic Hazard Map of India*, Working Committee of Experts (WCE), NDMA, New Delhi.
- Johnson R A and Wichern D W (2007) *Applied Multivariate Statistical Analysis*, Prentice-Hall, Upper Saddle River, New Jersey.
- Molina S, Lindholm C D and Bungum H (2001) "Probabilistic Seismic Hazard Analysis: Zoning Free versus Zoning Methodology." *Bollettino Di Geofisica Teorica Ed Applicata* **42(1-2)**: 19-39.
- Ordaz M, Augilar A and Arboleda J (2007) *CRISIS2007 Ver 1.1: Program for Computing Seismic Hazard*, Instituto de Ingenieria UNAM, Mexico.
- Ragunathan P (2011) *Development of Design Ground Motions Using Kernel Density Estimation Technique for Chennai and Koyna*, M.Tech. Thesis, Department of Civil Engineering, IIT Madras, Chennai.
- Silverman B W (1986) *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability, Chapman and Hall, London
- Stepp J C (1973) "Analysis of Completeness of the Earthquake Sample in the Puget Sound Area." NOAA Technical Report ERL 267-ESL 30 S. T. Harding (ed.) *Contributions to Seismic Zoning* 16-28.
- Vere-Jones D (1992). "Statistical Methods for the Description and Display of Earthquake Catalogs." *Statistics in the Environmental and Earth Sciences* A. T. Walden and P. Guttorp (eds.), Arnold Publishers, London, 220-246.

Woo G (1996) "Kernel Estimation Methods for Seismic Hazard Area Source Modeling." *Bulletin of the Seismological Society of America* **86(2)**: 353-362.

Wu T J, Chen C F and Chen H Y (2007) "A Variable Bandwidth Selector in Multivariate Kernel Density Estimation" *Statistics and Probability* **77**: 462-467.