Proceedings of International Conference on Advances in Civil Engineering
ACE-2011, 21-23 October 2011, INV pp. 32-46

Seismic Hazard Analysis for Chennai using Cluster-Based Kernel Density Estimation

P. Ragunathan
Formerly Post Graduate Student, Indian Institute of Technology Madras, Chennai – 600 036
C. K. Ramanna
Research Scholar, Indian Institute of Technology Madras, Chennai – 600 036
G. R. Dodagoudar
Associate Professor, Indian Institute of Technology Madras, Chennai – 600 036, goudar@iitm.ac.in

ABSTRACT: The probabilistic way of analyzing seismic hazard is reasonable in considering the seismicity distribution of all possible magnitudes and all possible distances over a specified time period. The conventional way of defining the seismicity by dividing the seismic source into zones for distributed seismicity, characterizing each zone with Gutenberg-Richter recurrence relationship and assuming uniform seismicity within each zone has certain drawbacks. In this study, zone free approach is proposed to evaluate the spatial distribution of seismicity based on non-parametric density estimation technique i.e. kernel density estimation technique (KDE). The fixed bandwidth kernel deals poorly with the earthquake distributions since the earthquake catalogue have several areas of high activity clusters and low background seismicity. Therefore in the present study, adaptive kernel technique using clustering method is proposed and applied to find the spatial activity rate density functions and integrated with other forms of uncertainty in magnitude and distance to determine the annual rate of exceedance of the selected ground motion parameter. In order to test the practical applicability of the proposed method, Chennai region (low to moderate seismicity region) has been chosen which lies in the intraplate seismicity region with diffused and distributed seismic characteristics. The results of the hazard analysis are provided in the form of peak ground acceleration (PGA) and uniform hazard spectra (UHS) for different return periods at the bed rock level. The PGA and UHS obtained from the present study for Chennai are compared with the Cornell-McGuire and the fixed kernel methods.

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

INTRODUCTION

Seismic hazard analysis (SHA) is used to estimate input motions in the form of strong ground motion parameters for a given site for the purpose of earthquake-resistant design of new structures and seismic safety evaluation of existing structures and/or facilities. The deterministic methodology basically aims at finding the combination of maximum possible magnitude and distance which would generate the highest level of ground motion at the site of interest. But the ground motion in different frequency range may be dominated by the earthquakes of different magnitudes and distances. It is necessary to consider the effects of all earthquakes of different magnitudes with their proper spatial distribution around the site of interest and not just a single earthquake[1]. The probabilistic way of finding seismic hazard was developed by Cornell[1] and McGuire and Arabasz[2] by introducing seismic sources in the seismogenic regions based on the regional seismotectonics and geologic setting. The works by Cornell[1], Lomnitz[3], Lomnitz and Rosenblueth[4], Reiter[5], Turcotte[6], Vere-Jones[7], Woo[8] and McGuire[9], among others, provide excellent reviews of seismic hazard analysis, and give exhaustive references on the topic of probabilistic seismic hazard analysis (PSHA).

The conventional way of defining seismicity in the hazard analysis has certain drawbacks like difficulty in delineating seismic sources into zones in the case of diffused and/or distributed seismicity, requirement of expertise knowledge, abrupt change in the seismicity at the zonal boundaries[10], applicability of Gutenberg-Richter (G-R) recurrence relation to low to moderate seismicity region is questionable, assumption of homogenous seismicity leads to bias in hazard value[11]. Therefore in the present study, an attempt is made in the area of seismic hazard analysis to find the spatial activity rate density functions using clustering based adaptive kernel density estimation technique. Then the seismic activity rate function is integrated with the other forms of uncertainty to determine the probability of exceedance of the selected ground motion parameter.

The kernel technique is a density estimation technique falls under the category of non-parametric method of determining the probability density function (PDF). In non-parametric density estimation technique, a prior assumption of the underlying density distribution is not made unlike the parametric density estimation techniques. Hence this form of density estimation technique is superior to the parametric technique. Since earthquakes are discrete and spatially continuous in their distribution, the spatial activity is modelled using *kernel density estimation technique.* The kernel density estimation (KDE) technique is more popular in many areas of earthquake engineering such as hazard analysis[8], occurrence rate determination[12], determination of influence area of an earthquake[13] and source size characterization to name a few. The KDE technique is suitable for data sets which are not zone-based but a point- or a line-based[14,15]. In the present study, the KDE technique has been used to determine the seismic activity rate density function which is used in characterizing the seismic sources in terms of seismic spatial activity rate. Further the technique is enhanced by the usage of clustering based adaptive kernel density estimation (AKDE) technique which is statistically superior to the normal or fixed kernel density estimation (FKDE) technique. The AKDE technique has the advantage in case of multimodal distributions and for smoothing the long tail distributions where undersmoothing in the tail regions is likely to cause difficulties[14].

The FKDE technique has already been applied to a few sites in South India such as Chennai[16] and Kanchipuram[17] where the hazard results matched well with the Cornell-McGuire approach. It has been observed by the researchers like Molina et al.[18] and Beauval et al.[11] that the FKDE technique yields same results for low to moderate seismicity region and lower hazard values in high seismicity regions when compared to the Cornell-McGuire approach. Ramanna and Dodagoudar[19] carried out the hazard analysis using AKDE technique for Chennai. They used the three step procedure of Silverman[14] for determining the variable bandwidth utilized in defining the spatial probability density surfaces and the remaining part of the hazard analysis was carried out as per Woo[8] methodology. They compared the peak ground acceleration (PGA) values of the fixed and adaptive kernel techniques and concluded that the adaptive kernel technique performs better than the fixed kernel technique for multimodal density estimation. In the present study, an attempt has been made to apply the AKDE technique which uses the clustering method to Chennai region and compares the results with the Cornell-McGuire approach and the fixed kernel technique. 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.

ADAPTIVE KERNEL DENSITY ESTIMATION

The kernel density estimation technique is a non-parametric way of finding the probability density for the data set which are non-zone based by placing density curves of any form (Triangular, Rectangle, Epanechnikov, Gaussian) on each data or sample point with some spread and summing the effect of these kernels[15] (Fig. 1). The measure of the spread of kernel is called *bandwidth* or *window width* which plays an important role rather than the type of kernel. Basically there are two techniques of kernel density estimation namely fixed bandwidth kernel and adaptive or variable bandwidth kernel techniques. The adaptive kernel techniques are of two types − if kernels are placed on the sample point, then the estimator is known as *sample point estimator* and if they are placed on the estimation points, it is called *balloon estimator*. In both cases, the density at any point is calculated as the normalized sum effect of each of these kernels. It is well known that, in case of seismic hazard analysis (SHA) the uncertainty lies in the epicenter (sample point) rather than the evaluation point (where the seismicity is estimated) and hence the sample point estimator is used in this study.

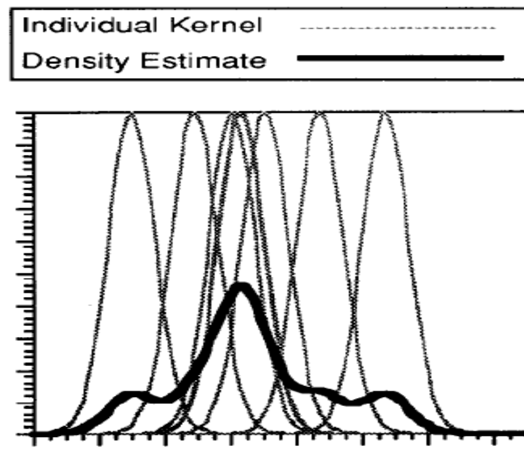

Fig. 1 Kernel density estimate

A multivariate adaptive kernel density function is of the form:

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

where n is the number of sample data $\mathbf{x}_i$, h_i is the variable bandwidth equal to b_iH, where b_i is the local bandwidth factor and H is the global bandwidth, $\mathbf{x}$ is the estimation or evaluation point and $K(.)$ is the kernel of any form e.g. a standard normal kernel such as

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

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

There are several techniques of determining the optimum bandwidth parameters for the dataset in hand, since choosing large bandwidth smoothens the density functions whereas smaller bandwidth makes the density functions too spiky (see Fig. 2). Therefore, there is a need to find an optimal bandwidth to be used for each data point to represent the exact relationship between earthquake catalogue and parent seismicity distribution in terms of spatial activity rate. In this study, the clustering based AKDE technique has been used to determine the spatial activity rate which is used in characterizing the seismic sources and hazard analysis.

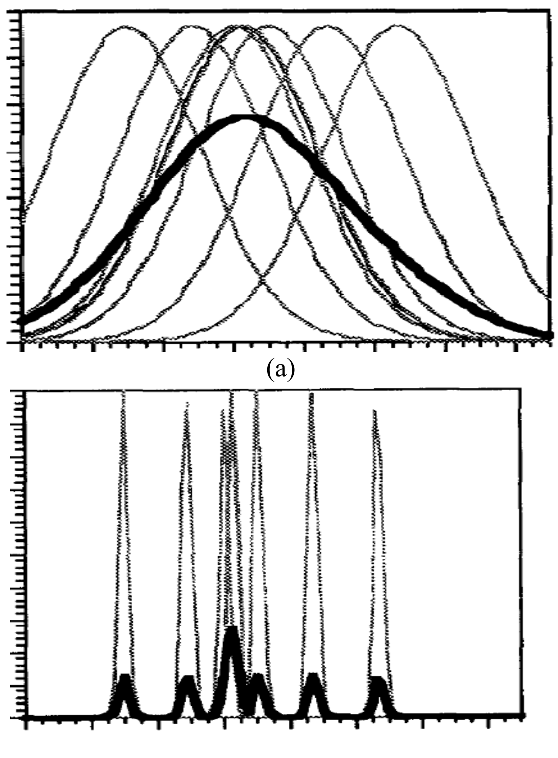

Fig. 2 Effect of bandwidth: (a) Larger bandwidth – smoothened PDF, (b) Smaller bandwidth – spiky PDF

Clustering Technique for Adaptive Bandwidth

Clustering or grouping analysis is a more primitive technique in that no assumptions are made concerning the number of groups or the group structure. Grouping is done on the basis of similarities or distances. Hierarchical clustering techniques proceed by either series of successive merges or a series of successive divisions. The earlier one is called agglomerative hierarchical clustering and the latter one is called divisive hierarchical clustering.

Zamani and Hashemi[20] used the clustering procedure for tectonic zoning of Iranian plateau. They pointed out that the hierarchical clustering can be used for seismic zoning and hazard estimation of seismogenic regions. Therefore the local bandwidth factor b_i used in determining h_i in Eq. (1) is determined by clustering method in which agglomerative hierarchical procedure is adopted[21]. 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.

Let the samples or observations be x_1, x_2,…, x_n. The algorithm to find the local bandwidth factors is as follows:[21]

1) Find the distance matrix $D = d_{ij}$ using n clusters or observations (initially) where d_{ij} is the $n \times n$ symmetric matrix.
2) Search the distance matrix for the nearest pair of clusters. Let the average distance between the nearest clusters U and V be

$$d_{UV} = \frac{1}{n_U n_V} \sum_{i=1}^{n_U} \sum_{j=1}^{n_V} d_{ij} \qquad (3)$$

where d_{ij} is the average distance between the observations i in the cluster U and j in the cluster V, n_U and n_V are the number of observations in the clusters U and V respectively.
3) Merge the clusters U and V. Label the newly formed cluster as UV. Update the entries in the distance matrix by deleting the rows and columns corresponding to the clusters U and V and by adding the rows and columns giving the distances between cluster UV and remaining clusters.
4) Repeat Steps 2 and 3 a total of n - 1 times so that all observations will be in a single cluster at the termination of algorithm. Record the identity of clusters that are merged and the distance levels at which the merges take place.
5) Let b_i is the average distance level of x_i in the dendrogram or local bandwidth factor

$$b_i = \frac{1}{n} \sum_{j=1}^{n_i} l_j \qquad (4)$$

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,… l_n are the distance levels at which n merges take place.

Evidently if the number of merges n is large then the local bandwidth factor b_i will be small which means that the observation or sample x_i located in the main part or dense region and eventually the corresponding density will be high and vice versa. A computer program in FORTRAN is developed for finding the local bandwidth factor b using clustering procedure and its reliability is tested against standard data. Figure 3 shows the illustration of the determination of local bandwidth factors.

SEISMIC HAZARD ANALYSIS: AKDE TECHNIQUE

The assessment of seismic hazard for a given site requires the evaluation of ground motion acceleration at that site. This can be determined by combining magnitude of earthquakes in the region with the attenuation of epicentral magnitude for a specified probability distribution of the distance from the epicenter. Therefore, one must determine the spatial distribution of epicentral distance for the earthquakes of the region, which is dictated by active faults or point sources if they are present in the region. The kernels have been used in various forms in the seismic hazard analysis. The very first researcher to use kernels in a vague sense was Bender[10] to handle the abrupt change of seismicity rate at zonal boundary where each point in the source zone was regarded as normally distributed. Woo[8] used the fixed kernel technique to replace the mean annual rate of exceedance υ_i for magnitude M of G-R recurrence law[22] with spatial activity rate $\upsilon(M, \mathbf{x})$. This technique has been applied to develop uniform hazard spectra (UHS) for Pyrenean region[23] and for an LNG plant at Taranto[24]. The kernel techniques have also been used for source site characterization in mining induced seismicity[25]. A modified version of Woo technique can be found in the work of Chan and Grunthal[26] known as hybrid method wherein the fixed kernel method was applied after forming source zones (based solely on geology) to arrive at uniform seismic hazard map for entire Europe. The first work on adaptive kernel application to earthquake engineering was by Stock and Smith[12,27].

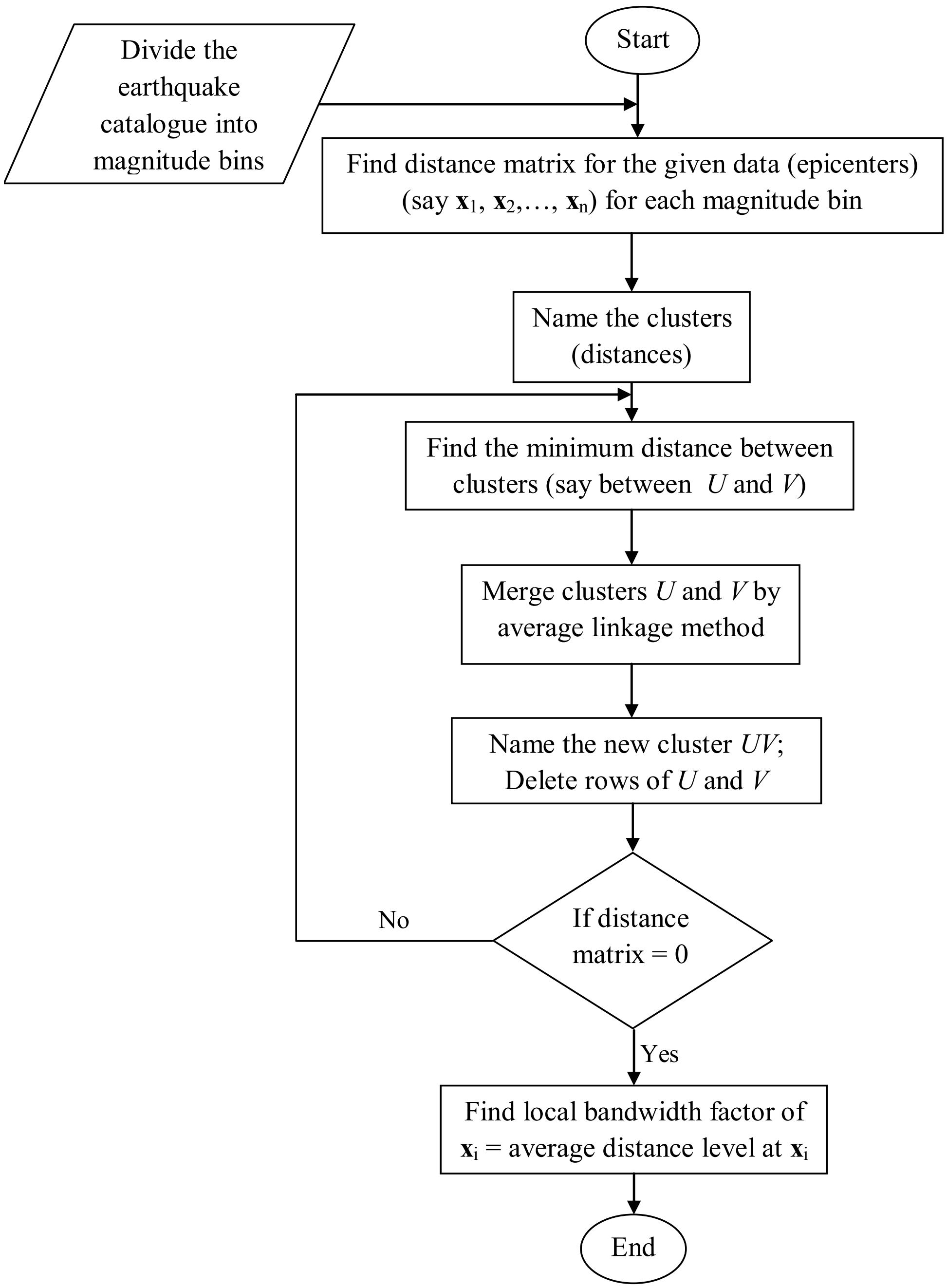

Fig. 3 Determination of local bandwidth factor using clustering technique

Earthquake occurrence in hazard analysis is often described as a homogenous stationary Poisson process or a Poisson cluster process. The Poisson model is memoryless, one could argue that this model is inappropriate to account for the occurrences of earthquakes since crustal deformations between major tectonic plates on well defined faults considered to have memory. However, the Poisson model can be applied in many situations where fault memory actually exists; it is inappropriate only if the elapsed time between significant events, with memory, exceeds the average recurrence time between such events[28]. The probability of ground motion parameter Y at a given site, exceeding a specified level y during a specified time t is given by

$$P(Y > y) = 1 - e^{-\lambda_{y*} t} \leq \lambda_{y*} t \tag{5}$$

where λ_{y*} is (mean annual rate of exceedance) the average frequency during time period t at which the level of ground motion parameter Y exceeds level y at a given site. The parameter λ_{y*} incorporates the uncertainty in time, size and location of future earthquakes and uncertainty in the level of ground motion they produce at the site. It is given as

$$\lambda_{y^*} = \sum_{i=1}^{N_S}\sum_{j=1}^{N_M}\sum_{k=1}^{N_R} \upsilon(M,\mathbf{x}) P\left[Y > y^* | m_j, r_k\right] P\left[M = m_j\right] P\left[R = r_k\right] \tag{6}$$

where $\upsilon(M,\mathbf{x})$ is the spatial activity rate density function, N_s is the number of potential earthquake sources, N_M is the range of magnitudes, N_R is all the possible range of distances from site to source, The $P\left[M = m_j\right]$ is included using magnitude error by smearing the activity rate. In conventional technique, $P\left[R = r_k\right]$ is determined by assuming that the earthquake is equally likely to occur at any point in the source but in the kernel method, it is replaced by incorporating the location error of epicenters. The $P\left[Y > y^* | m_j, r_k\right]$ is the probability that the ground motion level y^* will be exceeded for magnitude m_j and distance r_k. In the kernel technique, the seismicity rate υ_i is replaced by the spatial activity rate density function and is expressed as

$$\upsilon(M,\mathbf{x}) = \sum_{i=1}^{N} \frac{K(M, \mathbf{x} - \mathbf{x}_i)}{T_i} \tag{7}$$

where N is the number of earthquake events, $\mathbf{x}$ is the observation/estimation point, $\mathbf{x}_i$ is the epicentral location and T_i is the effective return period evaluated using the following expression:

$$T_i = \sum_i p_i D_i$$

where p_i being the detection probability of the event in a particular time period D_i and a numerical value is assigned to it based on the seismicity of the region (offshore or onshore). For example, the value of T_i for magnitude bin 4-4.49 is given in Tables 1 and 2.

Vere-jones[7] suggested the use of an anisotropic kernel for earthquakes and is given as

$$K(M,r) = \frac{n-1}{\pi h^2(M)} \frac{1+\delta \cos^2 \phi}{1+(\delta/2)} \left(1 + \left(\frac{r}{h(M)}\right)^2\right)^{-n} \tag{8}$$

where n is the exponent of the power law or kernel fractal scaling index taking value between 1.5 and 2. The δ is a measure of the strength of the directionality over isotropy - a value of zero implies isotropy, and a value of 10 or more imply significant anisotropy. Anisotropic kernel is useful when the activity rate associated with a particular fault needs to be determined. However in this study, isotropic kernel was used for Chennai city as it lies in the southern part of Peninsular India which is known for its distributed and diffused seismicity. The ϕ is the angle indicating the direction of an active lineament including the event epicenter. Such a lineament may be a specific geological structure, or an alignment of earthquake epicenters. The convention for defining the orientation of the lineament is that this angle is measured anti-clockwise from due east. Thus a lineament oriented due east would be associated with a zero angle, and a lineament oriented due west would be associated with an angle of 180 degrees. The variable bandwidth h is a function of magnitude and r (or $\mathbf{x} - \mathbf{x}_i$), is the distance to the epicenter. The above parameters are required in the kernel smoothing function and are called kernel parameters. It has been observed from the sensitivity study that the value of n has little effect on the results[19]. The value of n is taken as 1.75 in the present study The global bandwidth of the kernel for magnitude M is determined as

$$H(M) = ce^{(dM)} \tag{9}$$

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

The PSHA using AKDE technique consists of two parts. Firstly determine the activity rate υ(M,x) using adaptive kernel technique and secondly, integrate υ(M,x) with other forms of uncertainty to determine λy* given in Eq. (6) as per Woo methodology. The first part uses the clustering procedure and nearest neighborhood method for determining the bandwidth which is used in finding the spatial activity rate density function υ. In the adaptive kernel technique, h is a function of both the magnitude and space i.e. h(M, xi) hence K(·) varies spatially for each magnitude M. In the

TABLE 1 REFERENCE YEAR CALCULATION FOR MAGNITUDE BIN 4 - 4.49					
Onshore			Offshore		
Magnitude bin (4.0 - 4.49)			Magnitude bin (4.0 - 4.49)		
Time period	Probability (p_i)	Effective return period (p_iD_i)	Time period	Probability (p_i)	Effective return period (p_iD_i)
1500 - 1800	0.025	7.50	1500 - 1800	0.01	3.00
1800 - 1850	0.25	12.50	1800 - 1850	0.15	7.50
1850 - 1900	0.35	17.50	1850 - 1900	0.25	12.50
1900 - 1950	0.50	25.00	1900 - 1950	0.30	15.00
1950 - 1960	0.60	6.00	1950 - 1960	0.50	5.00
1960 - 1970	0.75	7.50	1960 - 1970	0.65	6.50
1970 - 1980	0.85	8.50	1970 - 1980	0.75	7.50
1980 - 1985	0.88	4.40	1980 - 1985	0.80	4.00
1985 - 1990	0.92	4.60	1985 - 1990	0.85	4.25
1990 - 1995	0.95	4.75	1990 - 1995	0.90	4.50
1995 - 2000	0.98	4.90	1995 - 2000	0.95	4.75
2000 - 2005	0.98	4.90	2000 - 2005	0.98	4.90
2005 - 2008	0.98	2.94	2005 - 2008	0.98	2.94
2008 - 2010	0.98	1.96	2008 - 2010	0.98	1.96
Total $\sum p_iD_i$		112.95	Total $\sum p_iD_i$		84.30
Reference year (current year - $\sum p_iD_i$)		1897.05	Reference year (current year - $\sum p_iD_i$)		1924.70

Note: In this study, current year is taken as 2010.

TABLE 2 REFERENCE YEAR FOR ALL MAGNITUDE BINS					
Onshore			Offshore		
Range	Reference year		Range	Reference year	
> 5.5	1861.59	□1□□□	> 5.5	1885.59	□1□□□
5.0 - 5.49	1873.78	□ 1874	5.0 - 5.49	1897.03	□ 1897
4.5 - 4.99	1885.38	□ 1885	4.5 - 4.99	1909.08	□ 1909
4.0 - 4.49	1897.05	□ 1897	4.0 - 4.49	1924.70	□ 1925
3.5 - 3.99	1909.08	□ 1909	3.5 - 3.99	1954.27	□ 1954
3.0 - 3.49	1921.18	□ 1921	3.0 - 3.49	1955.22	□ 1955

second part, the annual rate of exceedance is calculated by clubbing the spatial activity rate with uncertainties in magnitude, location and ground motion. The uncertainty in the magnitude and the epicenter of each event can be represented explicitly by Gaussian error distributions, which depend on the extent, accuracy and reliability of observation of events (accuracy of earthquake catalogue). The Gaussian errors for magnitude and distance compound the smoothing process. The contribution of each historical event to the activity rate is calculated using the kernel function. In ascertaining the contribution of each catalogued event to a particular activity rate density field, the estimated event magnitude is smeared over a normal distribution of values, with the prescribed standard deviation of magnitude error. Similarly, the estimated event location is smeared over a bivariate normal distribution of values, with the standard deviation of location error. The most recent is the earthquake, the lesser is the value of error and also, the smaller is the earthquake magnitude, the higher is the error. The location and magnitude errors for the Peninsular India earthquakes are assigned and used in the hazard analysis.

SEISMICITY OF CHENNAI

Chennai city is the fourth largest city located in the southern part of PI on the coromandel coast of Bay of Bengal which covers an area 172 km^2 and is located between 12.75°-13.25° N and 80.0°-80.5° E. The seismic map drawn by Bureau of Indian Standards in 2002 has shifted Chennai from Zone II (lower activity zone) to Zone III (moderate activity zone) due to increased seismicity (earthquakes like

Latur 1993, Jabalpur 1997, Bhuj 2001) of Peninsular India. Though it is in Zone III (low to moderate seismicity region), there are no major faults causing earthquakes in the region. But the surrounding regions have been identified with the faults which show evidence of movement during the Holocene period. The east-west trending Cauvery Fault, Tirukkavilur-Puducherry Fault and Vaigai River Fault and the north-south trending Comorin-Point Calimere Fault and Rajapatnam-Devipatnam Fault run close to major urban centers like Coimbatore, Madurai, Nagapattinam, Thanjavur and Puducherry. However, it must be stated that proximity to faults does not necessarily translate into a higher hazard as compared to areas located farther away, as damage from earthquakes depends on numerous factors such as subsurface geology and local site conditions. Very rare earthquakes of magnitude more than 5 with a maximum historic event of M_W = 6.0 which occurred in 1900 at Coimbatore[29], have been recorded. The earthquakes in this region are due to intraplate stresses within the preexisting weak zones.

Subrahmanya[30] observed a major compression lineament known as Mulki-Pulicat Axis (MPA) running from Mulki coast (approximately 13° N) on the east coast to Pulicat lake on the west coast (very close to Chennai city). This compression zone is the result of continuous spreading of sea floor in Indian ocean. The author found that the large number of small earthquakes in south of MPA whereas major earthquakes can be noticed only in North of MPA as it acts as a major block with significant stress accumulation. Recent study of the southern PI by Ramasamy[31] using remote sensing imageries has revealed several faults in the region. Seismically active regions like Ongole (15.60° N, 80.10° E) M_W = 5.2 in 1959 and 1969 and Coimbatore (10.8° N, 76.8° E) M_W = 6.0 in 1900 are being observed in Ongole-Tamil Nadu and Kerala lineaments respectively. Also a few major earthquakes around Pondicherry (11.93° N, 79.83° E) with maximum M_W = 5.6 in 1867 and the other off coast with M_W = 5.5 in 2001 have been observed.

A numerous small earthquakes of maximum M_W = 4.6 have been observed around Bangalore (12.97° N, 77.58° E) region. Also near Kanchipuram (12.83° N, 79.75° E) a pre-instrumental earthquake of magnitude 5.05 in 1823 has been observed. Hence the tectonic features in the Chennai region (e.g. Palar river fault, Tirukkavilur Pondicherry fault) are capable of producing moderate earthquakes that can cause structural damage to buildings and heritage structures which are not earthquake resistant. In the present study, the seismic hazard is evaluated at the geographical location which corresponds to 13.08° N and 80.28° E in Chennai city. Figure 4 shows the geologic and tectonic features of the study area.

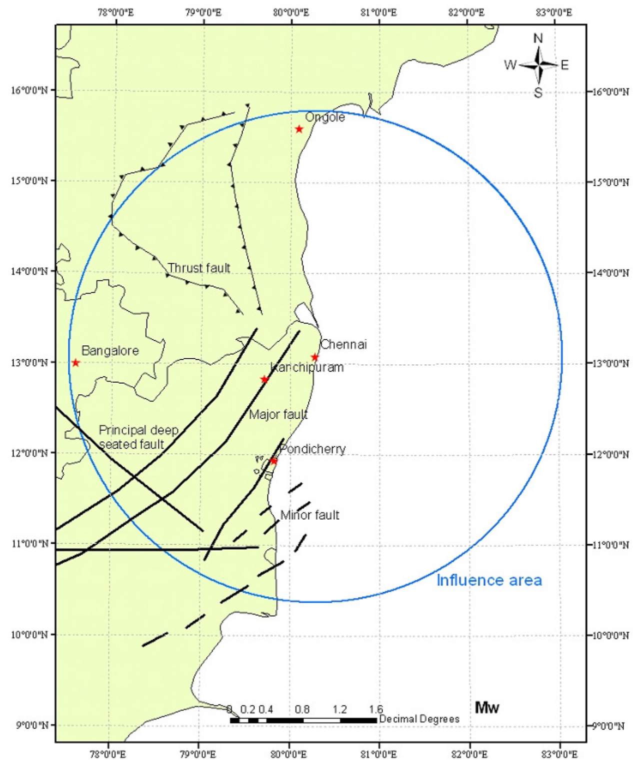

Fig. 4 Geologic and tectonic features of Chennai region (Gupta[29])

SEISMIC HAZARD ANALYSIS FOR CHENNAI – RESULTS

Nowadays the widespread availability of engineering work stations together with efficient solutions algorithm and sophisticated pre and post processing softwares enable the use of computational strong motion seismology (CSMS) in both research and engineering applications. The area of CSMS includes wave propagation, generation of synthetic strong motion data, elasto-dynamics, modelling of spatial variability of earthquake phenomenon, fracture process simulation in geologic media, modelling of earthquake source mechanics, etc. For the purpose of implementing the proposed method of clustering based AKDE technique to seismic hazard analysis, a computer program was developed in MATLAB which contains two parts. The first part of the program calculates the seismic activity rate density function at each grid location using kernel smoothing function and the second part combines the spatial activity rate with other forms of uncertainty to obtain the annual rate of exceedance of the selected ground motion parameter.

The total study area of 600 km × 600 km i.e. 300 km – control region around Chennai (13.08° N, 80.28° E) is divided into 10 km × 10 km grids. Previous studies on seismic hazard assessment carried out for Chennai[32-34,16,19] pointed out the need for better input data because of the low to moderate seismicity of the region. With this purpose, the present study is oriented towards the compilation of composite earthquake catalogue and accurate evaluation of spatial density function for the earthquakes of the region which in turn used in hazard computation for Chennai. The seismicity data required for the hazard analysis is collected from USGS, Chandra[35], India Meteorological Department (IMD), Gauri Bidanur Array (GBA) and National Earthquake Information Centre (NEIC). A total of 173 earthquakes of $M_w \geq 3.5$ are compiled for a circular area of 300 km radius around Chennai from the year 1507 to 2010 A.D. Fore and after shocks have been removed using Gardner and Knopoff[36] dynamic windowing technique which resulted in 151 main events. Thus obtained Poissoninan catalogue has been used for seismic hazard analysis. Figure 5 shows the statistics of composite catalogue before and after removal of the dependent events for Chennai region. The spatial distribution of Poissonian earthquake epicenters is shown in Fig. 6.

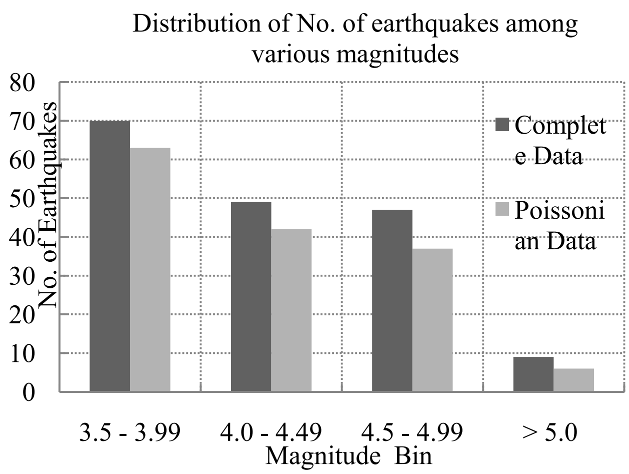

Fig. 5 Statistics of the composite catalogue for Chennai region

The kernel method of hazard analysis will not account for magnitudes greater than the historical maximum magnitude unless the uncertainties on 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[8]. The value of local bandwidth factor b is determined using clustering method. Figure. 7 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 (Fig. 8). The mean nearest distances for each magnitude bin are given in Table 3. In the kernel methodology of PSHA, the spatial activity rate $\upsilon(M,\mathbf{x})$ is the function of both 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 (Eq. 9).

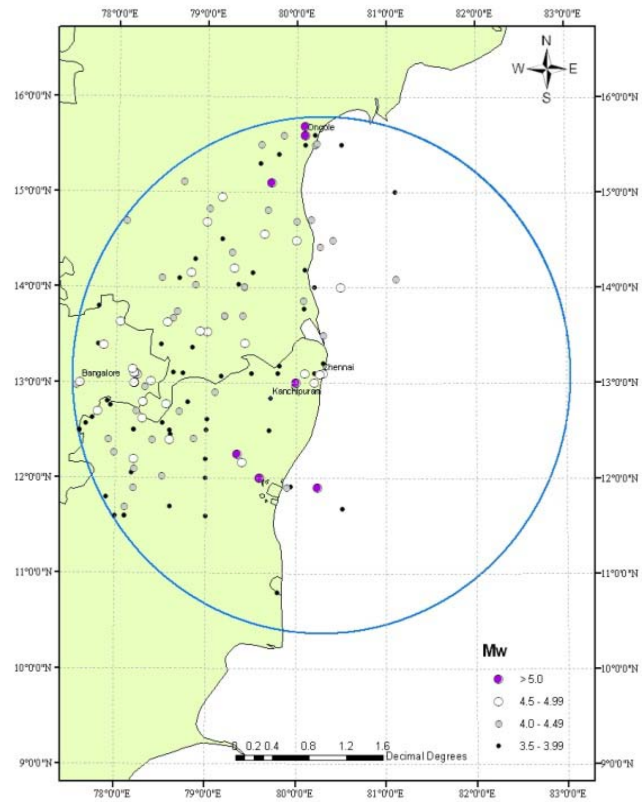

Fig. 6 Spatial distribution of declustered epicenters for Chennai

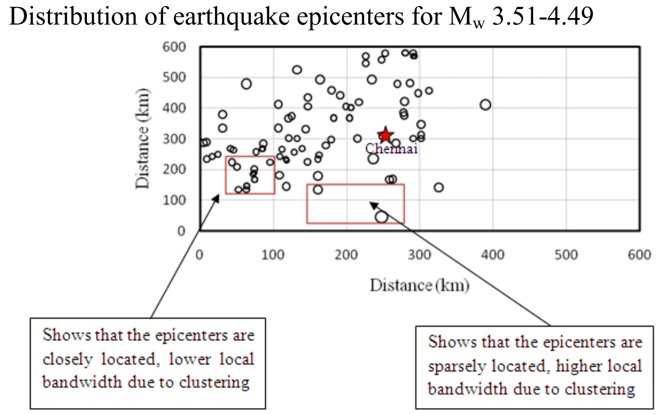

Fig. 7 Spatial distribution of epicenters for bandwidth determination

Determination of parameters c and d

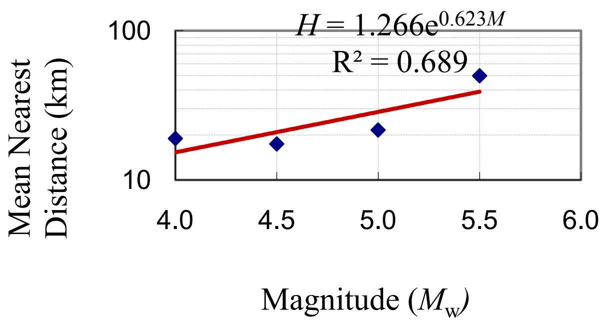

Fig. 8 Bandwidth parameters

TABLE 3
VALUES OF MEAN NEAREST DISTANCE FOR EACH MAGNITUDE BIN

Magnitude bin	Magnitude	Mean nearest distance
4.0 ± 0.49	4.0	18.98053
4.5 ± 0.49	4.5	17.42574
5.0 ± 0.49	5.0	21.59797
5.5 ± 0.49	5.5	49.92034

The attenuation relationship developed as part of the development of probabilistic seismic hazard map of India by Iyengar et al.[37] is used in the present study. The attenuation relationship was derived and validated for Himalayan region, Indo-Gangetic plains, NE region, Central India, Gujarat and southern India separately. This attenuation relationship accounts for geometrical spreading, inelastic attenuation and magnitude saturation similar to the finite source seismological model. The equation is

$$\ln(S_a/g) = c_1 + c_2M + c_3M^2 + c_4r + c_5\ln(r + c_6e^{c_7M}) + c_8\log(r)f_0 + \ln(\varepsilon) \quad (10)$$

where $f_0 = \max[\ln(r/100), 0]$, S_a is the spectral acceleration, M is the moment magnitude, r is the hypocentral distance in kilometers and c_1, c_2, c_3, c_4, c_5, c_6, c_7, and c_8 are the coefficients of attenuation equation whose values are tabulated for seven different regions separately in the technical report submitted to National Disaster Management Authority (NDMA), New Delhi. The predictions made by the attenuation relationship are compared with the three recorded strong ground motion data obtained from the Kutch after shock of M_w 5.7, 2001[38] (Table 4). From Fig. 9, it is clear that attenuation relationship suggested by Iyengar et al.[37] over predicts slightly the recorded motions for short distance earthquakes (such over prediction is well within the acceptable limits) and predicts well the recorded motion for larger distance earthquake.

The seismic activity rate is determined at each grid point and for each magnitude bin using the adaptive kernel technique considering Vere-Jones[7] isotropic kernel. The spatial activity rate is determined by forming magnitude bin with uncertainty of ±0.49. The hazard analysis using fixed kernel technique was carried out as per Woo methodology. Figures 10 and 14 show the spatial distribution of epicenters for magnitude bins 4.0 and 5.0 respectively. Figures 11 and 15 show the spatial activity rate of earthquakes using AKDE technique for magnitude bins 4.0 and 5.0 respectively. Similarly, Figs. 12 and 16 show the spatial activity rate of earthquakes using FKDE technique for magnitude bins 4.0 and 5.0 respectively. Figures 13 and 17 depict the difference in spatial probability surfaces for magnitude bins 4.0 and 5.0 respectively

TABLE 4
COMPARISON OF ATTENUATION RELATIONSHIP IN TERMS OF PGA WITH RECORDED DATA
(Singh et al.[38])

Station	Epicentral distance R (km)	PGA (g)		
		N	E	Z
28 January 2001 M_w 5.7 Kutch after shock at focal depth of 15 km				
BHUJ	101	0.0075	0.0079	0.0042
DGA	249	0.0017	0.0013	0.0011
BOM	576	0.0003	0.0002	0.0002

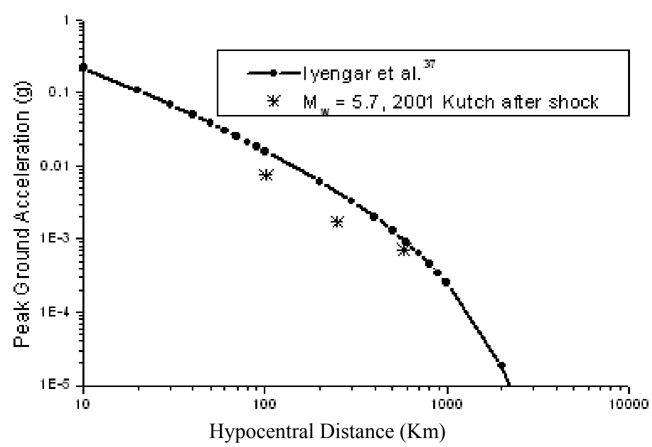

Fig. 9 Comparison of attenuation relationship in terms of recorded PGA

From Figs. 10 - 13, it is obvious that the spatial activity rate density function for lower magnitude bin resulted in multimodal distributions where the difference between two kernel techniques is explicitly visible at all epicenter locations. But the spatial activity rate density function for larger magnitude bin (Figs. 14 – 17) resulted in unimodal distribution due to few earthquakes of larger magnitude occurred around Chennai where the difference between two kernel techniques is visible only at the location of the peak value.

The so obtained seismic activity rate functions are clubbed with the other forms of uncertainties to obtain the probability of exceedance of the selected ground motion parameter. A hypocentral depth of 17 km for all epicenters was considered (Fig. 18) based on the seismicity data of Peninsular India. The variation of annual probability of exceedance with PGA by the clustering based AKDE technique is shown in Fig. 19

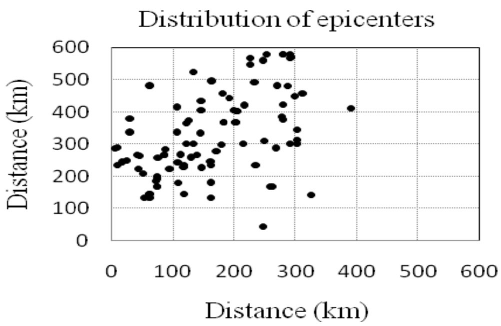

Fig. 10 Spatial distribution of epicenters for magnitude bin 4.0 (3.51 – 4.49)

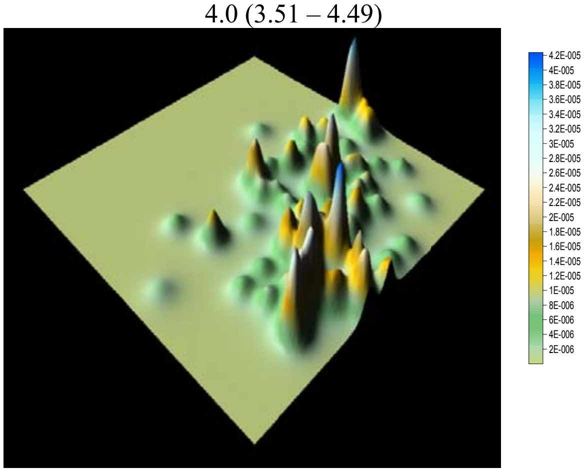

Fig. 11 Spatial activity rate for magnitude bin 4.0 from adaptive kernel technique

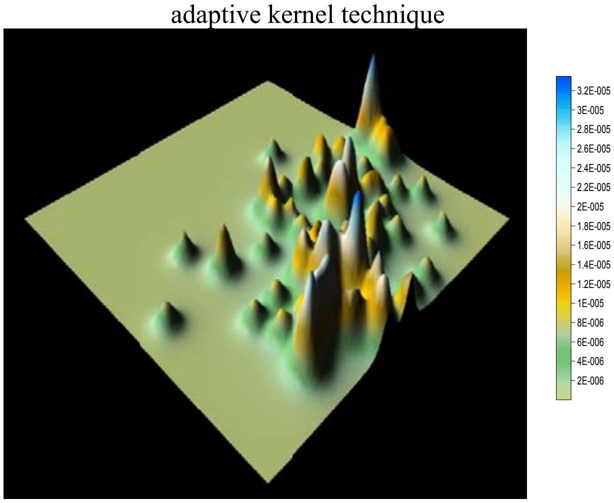

Fig. 12 Spatial activity rate for magnitude bin 4.0 from fixed kernel technique

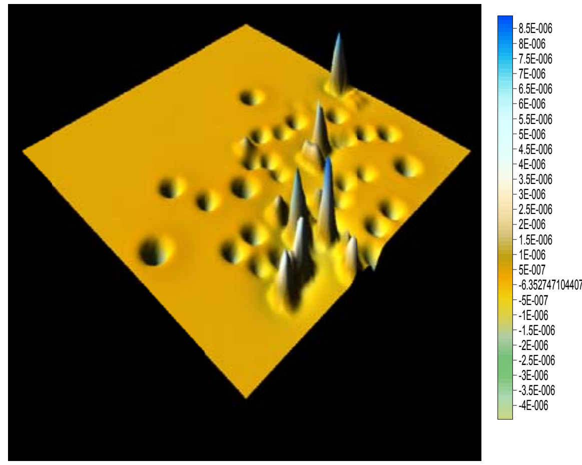

Fig.13 Difference in spatial probability surface for magnitude bin 4.0

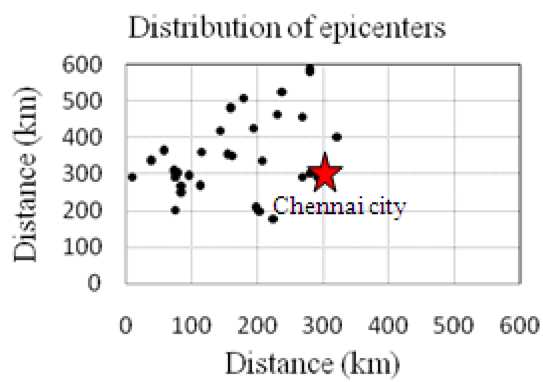

Fig. 14 Spatial distribution of epicenters for magnitude bin 5.0 (4.51 – 5.49)

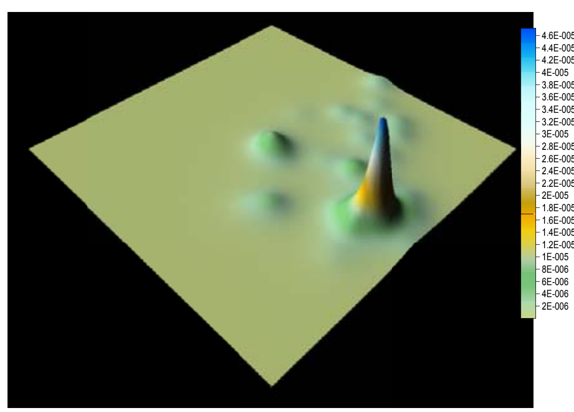

Fig. 15 Spatial activity rate for magnitude bin 5.0 from adaptive kernel technique

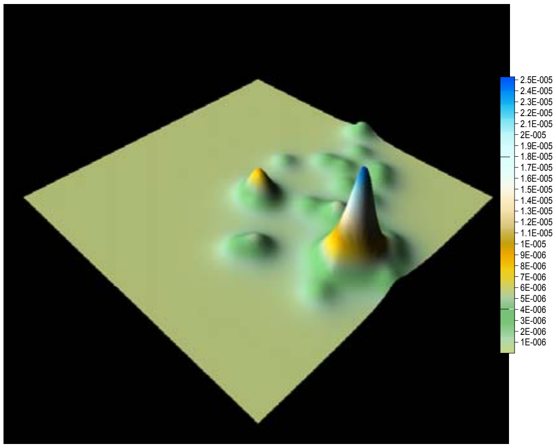

Fig. 16 Spatial activity rate for magnitude bin 5.0 from fixed kernel technique

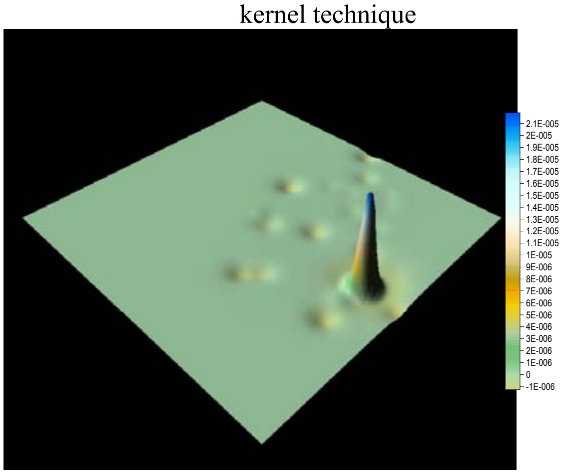

Fig. 17 Difference in spatial probability surface for magnitude bin 5.0

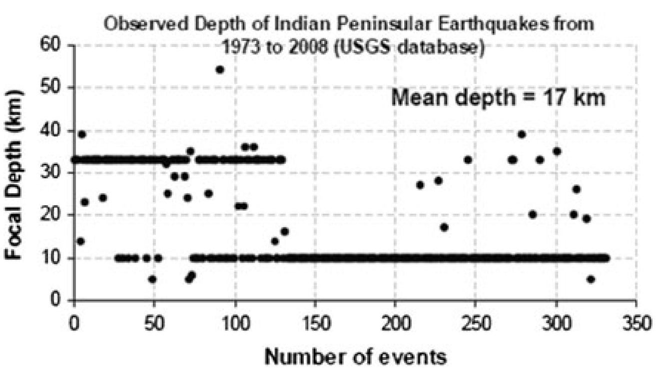

Fig. 18 Variation of focal depth for earthquakes in Peninsular India

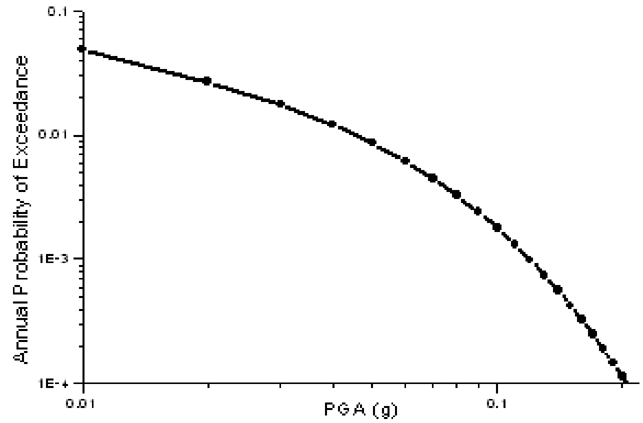

Fig. 19 Annual probability of exceedance by clustering based AKDE technique

CORNELL-MCGUIRE APPROACH

The core of conventional PSHA is the calculation of the probability of exceedance of a specified ground motion level at a particular site. This problem is decomposed into the calculation of the frequency of earthquakes with some damaging potential and the calculation of the conditional probability of exceedance for each of these contributing earthquakes, summed over all potentially contributing sources, that a specified ground motion level is exceeded. The methodology of PSHA using Cornell-McGuire approach is implemented using CRISIS 2007[39]. The catalogue completeness analysis was carried out using Stepp's method[40] considering a controlling area of 300 km radius as one single source zone (see Table 5 for completeness data). The G-R recurrence law parameters *a* and *b* are obtained (Fig. 20) and are given in Table 6. In this case also, the hazard analysis has been carried out using the above mentioned attenuation relationship.

The variation of annual probability of exceedance with PGA by the clustering based AKDE technique is shown in Fig. 21. Table 7 gives the values of PGA obtained from the AKDE technique and Cornell-McGuire method for different return periods. The comparative plots of UHS obtained by the AKDE technique, FKDE technique, Cornell-McGuire approach for 475 years return period and response spectra as given in IS 1893 (Part 1) 2002[41] for rock site are shown in Fig. 22. The response spectrum given in IS code corresponds to design basis earthquake (DBE) with 475 years return period. The PGA values predicted for 475 years return period by the proposed clustering based AKDE technique are compared with the values reported by the previous studies (Table 8).

TABLE 5 COMPLETENESS ANALYSIS		
Magnitude range	Completeness year	Completeness period (years)
3.5 - 3.99	1968	40
4.0 - 4.49	1968	40
4.5 - 4.99	1958	50
> 5.0	1800	209

TABLE 6 G-R RECCURENCE LAW PARAMETERS	
Parameter	Value
[illegible]	2.084
a	5.191
b	1.218
[illegible]	11.955
[illegible]	2.805

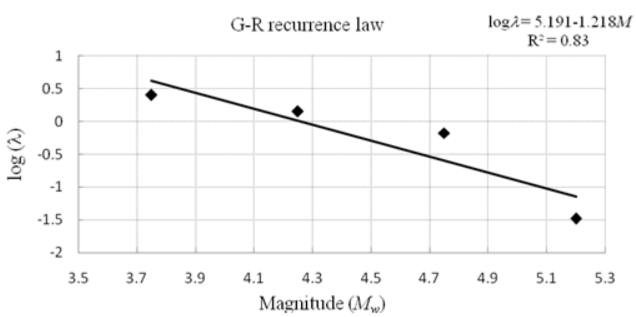

Fig. 20 G-R recurrence law for Chennai

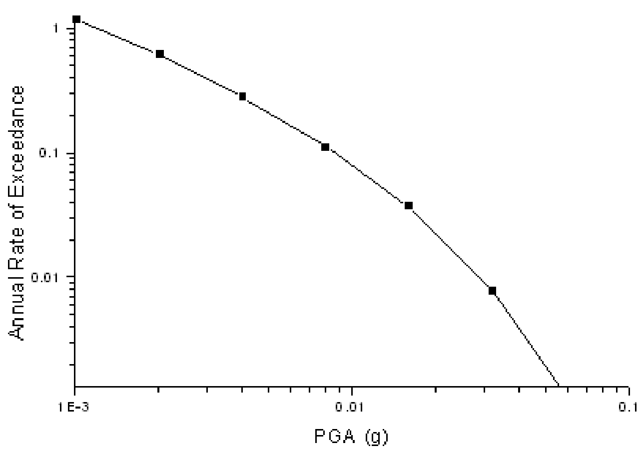

Fig. 21 Annual rate of exceedance by Cornell-McGuire approach

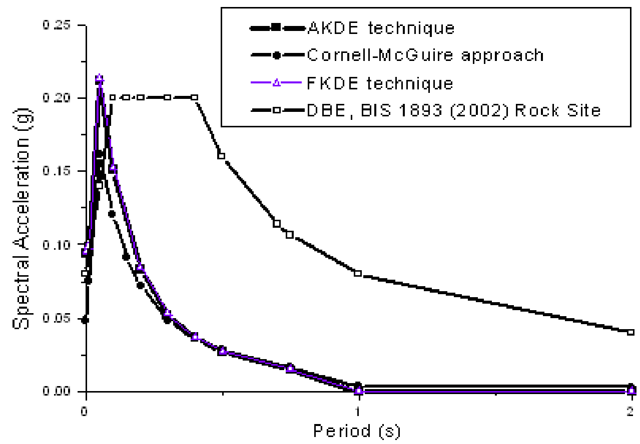

Fig. 22 Comparison of UHS for 475 years return period

TABLE 7 COMPARISON OF PGA VALUES FOR DIFFERENT RETURN PERIODS		
Return period (Years)	PGA (g) by Cornell-McGuire approach	PGA (g) by clustering based AKDE technique
72	0.025	0.036
100	0.028	0.043
475	0.048	0.094
975	0.060	0.118
2475	0.076	0.152

CONCLUSIONS

The clustering based adaptive kernel density estimation technique utilized the clustering procedure so that the denseness and sparseness nature of the earthquake epicenter distributions are clearly captured. The PGA values obtained for 475 years return period from the proposed clustering based AKDE technique and the previous studies are presented in Table 8. The AKDE technique resulted in the PGA value of 0.094g and is 17.5% more than the IS code specified value 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 (Fig. 22). From the figure it is noted that the AKDE technique yields higher value of hazard when compared to the Cornell-McGuire approach. But both the kernel methods yielded same PGA values of 0.094g. This can be attributed to the following: (1) Though the spatial activity rate density is higher for lower magnitudes in the clustering method, lower magnitude has less influence on the hazard estimation and hence there is no effect on the final hazard value. (2) Though the higher magnitudes have more influence on the hazard computation, Chennai being in low to moderate seismicity region, it resulted in unimodal distribution in density which is located far from the Chennai. For unimodal distributions, the

TABLE 8
PGA VALUES OBTAINED BY DIFFERENT APPROACHES FOR 475 YEARS RETURN PERIOD

Previous study	Approach	Catalogued period	Attenuation relationship used	PGA (g)
IS 1893: 2002[41]	–	–	–	0.08
Jaiswal and Sinha[42]	Frankel method of zoneless approach	1842 -2002	Iyengar and Raghu Kanth[43]; Atkinson and Boore[44]; Toro et al.[45]	0.06
Iyengar et al.[37]	Conventional method	2474 BC - 2008	Iyengar et al.[37]	0.065
Menon et al.[34]	Both zoning and fixed kernel	1063 -2008	Raghu Kanth and Iyengar[46]; Campbell and Bozorgina[47]; Abrahamson and Silva[48]	0.09
Ramanna and Dodagoudar[16]	Fixed kernel method	1507 -2008	Raghu Kanth and Iyengar[46]	0.083
Ramanna and Dodagoudar[19]	Adaptive kernel method (Silverman, 1986)	1507 -2008	Raghu Kanth and Iyengar[46]	0.087
Cornell-McGuire approach (Present study)	300 km radius around Chennai as a single zone	1507 -2008	Iyengar et al.[37]	0.05
Present study	Clustering based AKDE technique	1507 -2008	Iyengar et al.[37]	0.094

density obtained from the adaptive kernel technique is same and hence no change in the hazard value across the kernel techniques. It is concluded that though the clustering based AKDE technique has the superiority in exactly representing the seismic activity rate density functions and needs to be tested for moderate and high seismicity regions and for very low seismicity regions in south India.

For low to moderate seismicity regions where the earthquake data are scarce, it is difficult to carry out the completeness analysis as it induces subjectivity. In the conventional PSHA (i.e. Cornell-McGuire approach), the accuracy of the hazard calculations depends very much on the correctness of the completeness analysis results. In the kernel techniques, subjectivity is introduced owing to the detection probability p_i which needs to be assigned for every earthquake event. However this can be overcome by careful study of the paleoseismicity and instrumental seismicity of the region and a standard detection probability chart can be arrived for the region. The studies performed as part of the present investigation clearly indicate that the site specific hazard estimates should be considered together with code specified values, to arrive at safe and rational decisions concerning the selection of DBE for structures in Chennai region. The present kernel based method is consistent with the fractal nature of the earthquake process.

REFERENCES

[1] Cornell, C. A, "Engineering Seismic Risk Analysis", *Bulletin of the Seismological Society of America*, Vol. 58, 1968, pp. 1583-1606.
[2] McGuire, R. K. and Arabasz, W. J., "An Introduction to Probabilistic Seismic Hazard Analysis", *Geotechnical and Environmental Geophysics*, S. H. Ward (Ed.), Society of Exploration Geophysicists, Vol. 1, 1990, pp. 333–353.
[3] Lomintz, C., "*Global Tectonics and Earthquake Risk*", Elsevier, Amsterdam, 1974.
[4] Lomnitz, C. and Rosenblueth, E., "*Seismic Risk and Engineering Decisions*", Elsevier, Amsterdam, 1976.
[5] Reiter, L., "*Earthquake Hazard Analysis: Issues and Insights*", Columbia University Press, New York, 1990.
[6] Turcotte, D. L., "*Fractals and Chaos in Geology and Geophysics*", Cambridge University, Press. Cambridge, 1992.
[7] Vere-Jones, D., "Statistical Methods for the Description and Display of Earthquake Catalogs", A. T. Walden and P. Guttorp, (Eds.), *Statistics in the Environmental and Earth Sciences*, Arnold Publishers, London, 1992, pp. 220-246.
[8] Woo, G., "Kernel Estimation Methods for Seismic Hazard Area Source Modelling", *Bulletin of the Seismological Society of America*, Vol. 86, 1996, pp. 353–362.
[9] McGuire, R. K., *Seismic Hazard Risk Analysis*, Earthquake Engineering Research Institute, Berkeley, 2004.
[10] Bender, B., "Modelling Source Zone Boundary Uncertainty in Seismic Hazard Analysis", *Bulletin of the Seismological Society of America*, Vol. 76(2), 1986, pp. 329-341.
[11] Beauval, C., Scotti, O. and Bonilla, F., "The Role of Seismicity Models in Probabilistic Seismic Hazard Estimation: Comparison of a Zoning and a Smoothing Approach", *Geophysics Journal International*, Vol. 165, 2006, pp. 584-595.
[12] Stock, C. and Smith, E. G. C., "Adaptive Kernel Estimation and Continuous Probability Representation

of Historical Earthquake Catalogs", *Bulletin of the Seismological Society of America,* Vol. 92(3), 2002a, pp. 904-912.

[13] Chen, Y., Liu, J., Chen, L., Chen, Q. and Chan, L.S., "Global Seismic Hazard Assessment based on Area Source Model and Seismicity Data", *Natural Hazards*, Vol. 17, 1998, pp. 251-267.

[14] Silverman, B. W., "*Density Estimation for Statistics and Data Analysis*", Monographs on Statistics and Applied Probability, Chapman and Hall, London, 1986.

[15] de Smith, M. J., Goodchild, M. F. and Longley, P. A., "*Geospatial Analysis – A Comprehensive Guide*", 3rd ed., The Winchelsea Press, Leicester, 2009.

[16] Ramanna, C. K. and Dodagoudar, G. R., "Probabilistic Seismic Hazard Analysis Using Kernel Density Estimation Technique for Chennai, India", *GEORISK,* 2011a, 10.1080/17499518.2010.496073.

[17] Lai, C. G., Menon, A., Ornthammarath, T., Corigliano, M., Lizarraga, H. S. and Dodagoudar G. R., "*Probabilistic Seismic Hazard Assessment and Stochastic Site Response Analysis at the Archaeological Site of Kancheepuram in Southern India*", Research Report No. EUCENTRE-2009/01, University of Pavia, Pavia, Italy, 2009.

[18] Molina, S., Lindholm, C. D. and Bungum, H., "Probabilistic Seismic Hazard Analysis: Zoning Free Versus Zoning Methodology", *Bollettino Di Geofisica Teorica Ed Applicata*, Vol. 42(1-2), 2001, pp. 19-39.

[19] Ramanna, C. K. and Dodagoudar, G. R., "Seismic Hazard Analysis Using Adaptive Kernel Density Estimation Technique for Chennai City", *Pure and Applied Geophysics*, 2011b, DOI 10.1007/s00024-011-0264-8.

[20] Zamani, A. and Hashemi, N. "Computer-based Self-Organised Tectonic Zoning: A Tentative Pattern Recognition for Iran", *Computers and Geosciences*, Vol. 30, pp. 2004, 705-718.

[21] Johnson, R. A. and Wichern, D. W., "*Applied Multivariate Statistical Analysis*", Prentice-Hall, Upper Saddle River, New Jersey, 2007.

[22] Gutenberg, B. and Richter, C. F., "*Seismicity of the Earth and Related Phenomena*", Princeton University Press, Princeton, New Jersey, 1954.

[23] Secanell, R., Bertil, D., Martin, C., Goula, X., Susagna, T., Tapia, M., Dominique, P., Carbon, D. and Fleta, J., "Probabilistic Seismic Hazard Assessment of the Pyrenean Region", *Journal of Seismology*, Vol. 12, 2008, pp. 3-341.

[24] Principia., "*Seismic Hazard Evaluation, LNG Plant at Taranto (Italy)*", Report No.: 673 Rev. 1, Project No. P-373, 2005.

[25] Stainslaw, L. and Sikora, O. B., "Seismic Hazard Assessment under Complex Source Size Distribution of Mining - Induced Seismicity", *Tectonophysics*, Vol. 456, 2008, pp. 28-37.

[26] Chan, C. H. and Grunthal, G., "*Developing New Approaches to Earthquake Hazard Assessment and Forecasting*", Network for Research Infrastructures for European Seismology (NERIES), EC Project Number: 026130, Europe, 2008.

[27] Stock, C. and Smith, E. G. C., "Comparison of Seismicity Models Generated by Different Kernel Estimations", *Bulletin of the Seismological Society of America,* Vol. 92(3), 2002b, pp. 913-922.

[28] Cornell, C. A. and Winterstein, S. R., "Temporal and Magnitude Dependence in Earthquake Recurrence Models", *Bulletin of the Seismological Society of America,* Vol. 78, 1988, pp. 1522-1537.

[29] Gupta, I. D., "Delineation of Probable Seismic Sources in India and Neighborhood by a Comprehensive Analysis of Seismotectonic Characteristics of the Region", *Soil Dynamics and Earthquake Engineering*, Vol. 26, 2006, pp. 766–790.

[30] Subrahmanya, K. R., "Active Intraplate Deformation in South India", *Tectonophysics*, Vol. 262, 1996, pp. 231-241.

[31] Ramasamy, S. M., "Remote Sensing and Active Tectonics of South India", *International Journal of Remote Sensing*, Vol. 27, 2006, pp. 4397-4431.

[32] Sreedhar, U., "Studies on Probabilistic Seismic Hazard Assessment and Ground Response", *M. S. Thesis*, Department of Civil Engineering, IIT Madras, Chennai, 2007.

[33] Kalyan Kumar, G., Sreedhar, U. and Dodagoudar, G. R., "Probabilistic Seismic Hazard Analysis for Low Seismicity Region", *Indian Geotechnical Journal*, Vol. 39, 2009, pp. 288 -316.

[34] Menon, A., Ornthammarath, T., Corigliano, M. and Lai, C.G., "Probabilistic Seismic Hazard Macrozonation of Tamil Nadu in Southern India", *Bulletin of the Seismological Society of America*, Vol. 100, 2010, pp. 1320-1341.

[35] Chandra, U., "Earthquakes of Peninsular India, A Seismotectonic Study", *Bulletin of the Seismological Society of America*, Vol. 65, 1977, pp. 1387-1413.

[36] Gardner, J. K. and Knopoff, L., "Is the Sequence of Earthquakes in Southern California, with Aftershocks Removed, Poissonian?", *Bulletin of the Seismological Society of America*, Vol. 64, 1974, pp. 1363 – 1367.

[37] Iyengar, R. N., Chadha, R. K., Balaji Rao, K. and Raghu Kanth, S. T. G., "*Development of Probabilistic Seismic Hazard Map of India*", Working Committee of Experts (WCE), NDMA, New Delhi, 2010.

[38] Singh, S. K., Bansal, B. K., Bhattacharya, S. N., Pacheco, J. F. and Dattatrayam, R. S., "Estimation of Ground Motion for Bhuj (26 January 2001; M_w 7.6) and for Future Earthquakes in India", *Bulletin of the Seismological Society of America*, Vol. 93(1), 2003, pp. 353–370.

[39] Ordaz, M., Aguilar, A. and Arboleda, J., "*CRISIS2007 – Ver. 1.1: Program for Computing Seismic Hazard*", Instituto de Ingenieria, UNAM, Mexico, 2007.

[40] Stepp, J. C., "Analysis of Completeness of the Earthquake Sample in the Puget Sound Area", S. T. Harding, (Ed.), *Contributions to Seismic Zoning*, NOAA Technical Report ERL 267-ESL 30, 1974, pp. 16-28.

[41] IS 1893 (Part 1), "*Criteria for Earthquake Resistance Design of Structures*, *Part-1: General Provisions and Buildings*", Fifth Revision, Bureau of Indian Standards (BIS), New Delhi, 2002.

[42] Jaiswal, K. and Sinha, R., "Probabilistic Seismic-Hazard Estimation for Peninsular India", *Bulletin of the Seismological Society of America*, Vol. 97, 2007, pp. 318-330.

[43] Iyengar, R. N. and Raghu Kanth, S. T. G., "Attenuation of Strong Ground Motion in Peninsular India", *Seismological Research Letters*, Vol. 75, 2004, pp.530–540.

[44] Atkinson, G. M. and Boore, D. M., "Ground Motion Relations for Eastern North America", *Bulletin of the Seismological Society of America*, Vol.85, 1995, pp.17–30.

[45] Toro, G. N., Abrahamson, N. A. and Schneider, J., "Model of Strong Ground Motions from Earthquakes in Central and Eastern North America: Best Estimates and Uncertainties", *Seismological Research Letters*, Vol. 68, 1997, pp. 41–57.

[46] Raghu Kanth, S. T. G. and Iyengar, R. N., "Estimation of Seismic Spectral Acceleration in Peninsular India", *Journal of Earth System Sciences*, Vol. 116, 2007, pp. 199–214.

[47] Campbell, K. W. and Bozorgina, Y., "NGA Ground Motion Model for the Geometric Mean Horizontal Component of PGA, PGV, PGD and 5% Damped Linear Elastic Response Spectra for Periods Ranging from 0.01 to 10 s", *Earthquake Spectra*, Vol. 24(1), 2008, pp. 139-171.

[48] Abrahamson, N. A. and Silva, W. J., "Empirical Response Spectral Attenuation Relations for Shallow Crustal Earthquakes", *Seismological Research Letters*, Vol. 68(1), 1997, pp. 94–127.