# Regional level forecasting of seismic energy release

**B. Kavitha[1] · S. T. G. Raghukanth[1]**

**Abstract** This article explores a new strategy for forecasting of earthquake energy release in the seismogenic zones of the world. A total of 41 active seismogenic zones are identified with the help of past seismicity data. The magnitudes of individual events occurred in each zone are converted into seismic energy using an empirical relation. The annual earthquake energy time series is constructed by adding the energy releases of all the events in a particular year. The technique of principal component analysis is employed for the regionalization of these seismogenic zones using seismic energy time series. The annual energy time series of seismogenic zones are decomposed into finite number of intrinsic mode functions (IMFs) using ensemble empirical mode decomposition technique. The periodicities of the IMFs and their contribution to the total variance of the earthquake energy release are examined. The artificial neural network technique is used for modeling and forecasting the energy-time series of seismogenic zones. The model is verified with an independent subset of data and validated using statistical parameters. The forecast of the annual earthquake energy release in each seismogenic zone is provided for the year 2015.

**Keywords** Earthquake forecasting · Seismic energy time series · Principal component analysis · Empirical mode decomposition · Artificial neural network

## 1 Introduction

Earthquake prediction is an active topic in among researchers. As a complex physical phenomenon, it is very intricate to predict the magnitude and location of future event. Although epicentre of earthquakes seems to be random, the occurrence of earthquakes is more in the plate boundaries. The plate intersections are experiencing more number of earthquakes which are known as interplate events (Bolt 2005). But some of the plate

✉ S. T. G. Raghukanth
   raghukanth@iitm.ac.in

[1] Department of Civil Engineering, IIT Madras, Chennai, India

boundaries have more number of earthquakes than other boundaries as an indication of high seismic activity. These intersections are experiencing most of the large earthquakes ($M_w \geq 7$) and deeper earthquakes. These regions are called as seismogenic zones of the world. Hence it is necessary to study the seismicity pattern of these regions and it will of interest to forecast the future earthquake occurrence in these seismogenic regions.

The seismogenic regions with correlation among themselves can be grouped together. An advantage of such a grouping is that the time series data of seismogenic zones of same group, instead of individual zone data, could be used for effective forecasting. This grouping clearly reduces the size of the data to be handled in forecasting. Additionally, it also enhances some of the signals present on larger spatial scales (Nicholson, 1986). Identification of homogeneous zones, also called regionalization, can be attempted with the yearly or monthly data. The criterion adopted for delineating the groups can be based on variety of measures, such as means, coefficients of variation and correlation coefficients (CC). In the present study, principal component analysis (PCA) is used for the regionalization of global seismogenic zones using annual seismic energy time series. Principal components analysis is a multivariate statistical technique used to find a few mutually orthogonal linear combinations of the original variables which capture most of the variability present in the data. It is possible to capture the large part of the variance with a small number of components. This methodology has been widely used in meteorology (Kutzbach 1967; Overland and Preisendorfer 1982; Ehrendorfer 1987). Iyengar and Basak (1994) have used PCA technique to group the regions with homogeneous variability of monsoon rainfall.

In recent decades, investigating the seismic activities on global level has improved significantly because of the global seismograph networks. It is physically meaningful to express earthquake size in terms of released seismic energy than magnitude. Tsapanos and Liritzis (1992) studied the correlation in seismic energy release of three seismic regions namely Chile, Kamchatka and Mexico. Tsapanos (1998) evaluated the seismic hazard for eleven regions of the world using the released strain energy. Recently, Varga et al. (2011) analyzed the declustered catalogue of large earthquakes ($M \geq 7.0$) occurred from 1960 to 2011 and reported that the latitude distribution of the released seismic energy is bimodal with respect to the equator. Since the PCA technique is promising technique, it can be used to study the seismicity pattern of seismogenic zones. Telesca et al. (2004) have applied PCA approach to analyze the short-term variability of geoelectrical field measured at southern Italy. This analysis has proven that the PCA approach can be used for monitoring seismic areas. An approach based on the PCA has been used to study the distinctive characteristics of few earthquakes by Srivastava and Bhattacharya (1998). Differences in the mechanism of earthquakes were broadly supported by their seismicity patterns. Kavitha and Raghukanth (2015) have investigated the seismic energy release by considering the whole globe as a single unit. The regional level predictions are in great demand among common public, government agencies and policy makers due to their significance. Hence it is important to explore the seismic activities in a smaller spatial scale.

In this study, an effort is taken to understand the seismicity pattern of seismogenic zones and to forecast the earthquake occurrence in these seismogenic zones with the help of PCA technique. The ISC–GEM Global Instrumental Reference Earthquake Catalogue (http://www.isc.ac.uk/iscgem/index.php) has been utilized in the present study. A total of 41 active seismogenic zones are identified from tectonic plate boundaries using of past seismicity data. The events occurred in a particular seismogenic zone are separated from the main catalogue and thus a separate catalogue is prepared for each seismogenic zone. Choy and Boatwright (1995) empirical relation is used to convert the magnitudes of individual events in the catalogue of seismogenic zones into released seismic energy. The

annual seismic energy time series of seismogenic zones are constructed by adding the energy releases of all the events in a particular year. The empirical mode decomposition technique is used to extract finite number of intrinsic mode functions (IMFs) from the seismic energy time series. The influence of solar and lunar cycles on earthquake occurrence of each seismogenic zone is confirmed by the estimated periodicities and percentage variances of IMFs. In the present article, PCA is used for the regionalization of global seismogenic regions. The principal components are estimated from the annual seismic energy series of 41 seismogenic zones. The significance of principal components is estimated by the methodology proposed in detail by Preisendorfer et al. (1981). The significant principal components which have important signals of seismic energy release are identified in the present study. It helps to arrange zones in relation to their connection with these signals. In turn, this automatically ranks the regions also in terms of their importance to the annual seismic energy release. The derived PCs are used to organize 41 seismogenic zones into 16 groups. Due to the complexities involved, the auto regressive models are used for modeling and forecasting the earthquake energy data. The $IMF_1$ and the remaining part of the data are modeled separately by artificial neural network (ANN). The time series data of all seismogenic zones of a group, instead of individual zone data, is used for effective modeling. The proposed model is also validated using data which is excluded in the modeling period. The efficacy of the model has been verified by three statistical parameters, the root mean square (RMS) error $\sigma(\varepsilon)$, CC and performance parameter (PP). It is found that the developed model is efficient in forecasting the annual earthquake energy release of the seismogenic zones with a known error band.

## 2 Seismicity data

The first step of earthquake forecasting is the collection of good quality seismicity data with a long span of period. In the year 2013, the International Seismological Centre (http://www.isc.ac.uk/) has released the ISC–GEM Global Instrumental Reference Earthquake Catalogue spanning from 1900 to 2009. This catalogue consists of 18,809 events with magnitudes $M_W \geq 5.5$ all over the world. Events are reported in Moment magnitude scale ($M_W$) to maintain consistency. In the present article, this catalogue is updated till December 2014 using the recent seismicity data collected from ISC seismograph network. To estimate the exact amount of seismic energy released from a region, it is important to consider all the events occurred in that particular region. Hence the events with small magnitudes are also included in the catalogue. The events with different magnitude scales are converted into Moment magnitude using the same empirical relation used by ISC–GEM catalogue. The updated catalogue contains a total of 4,893,609 events with magnitude $M_W \geq 1$ spanning from 1900 to 2014. The events occurred before 1950 are omitted because there is no complete information about small magnitude events. The final catalogue has 4,880,684 events ($M_W \geq 1$) spanning from 1950 to 2014.

In the Fig. 1, the final catalogue is shown as a function of magnitude and time to ensure the magnitude completeness. Though the catalogue has short span (65 years), it is complete and homogenous. The events with all magnitude range are covered with reasonable increments from 1950 to 2014. Although there are some gaps from 1950 to 1960, it can be attributed to the lack of instruments in that period. There are 1108 events with magnitude Mw $\geq$ 7 and 61 events with magnitude Mw $\geq$ 8. The final catalogue has 5 events with magnitudes Mw $\geq$ 9. The '1960 Valdivia earthquake' with magnitude $M_w$ 9.6 is the largest
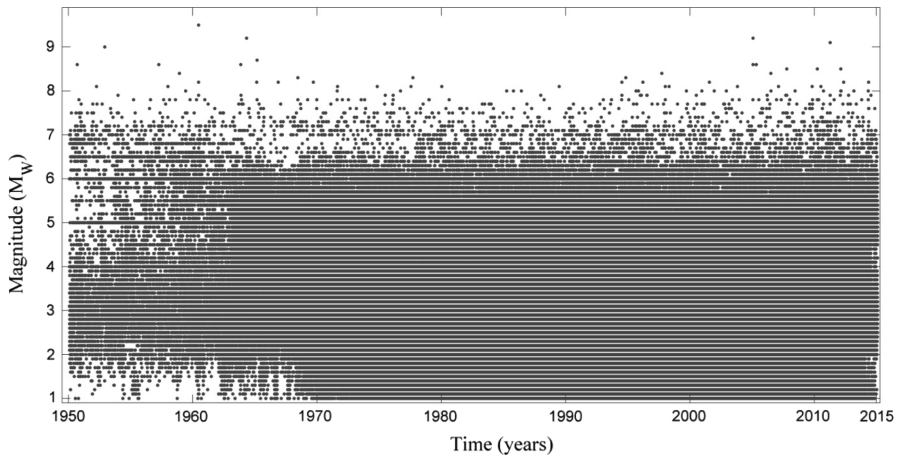
**Fig. 1** Final Catalogue spanning from 1950 to 2014

earthquake of the catalogue. The 1952 Kamchatka earthquake ($M_W$ 9.0), 1964 Alaska earthquake ($M_W$ 9.2), 2004 Sumatra–Andaman earthquake ($M_W$ 9.2) and 2011 Japan Earthquake of magnitude $M_W$ 9.1 are other large events in the catalogue. The catalogue consist of 90.7 % of shallow events (<70 km), 8.2 % of intermediate events (70–300 km) and around 1 % of deep events (>300 km).

Bird (2003) has developed the boundaries of present tectonic plates on Earth. The digital boundaries of these 52 plates (14 major and 32 minor plates) were presented in this model. By superimposing the epicenters of earthquakes on these plate boundaries, it is clear that seismic activity in the plate intersections is much higher than interior of tectonic plates (Fig. 2). Most of the large earthquakes (Mw ≥ 7) were occurred on these intersections only. Hence it will be interesting to study the seismicity pattern of these plate intersections. Although Bird (2003) demarcated 52 plate boundaries, some of them have very low seismic activity. So it is important to identify the plate boundaries with high seismic activity. In the present article, the intersection between any two tectonic plates is considered as the 'seismogenic zone'. For example, the intersection between Indian plate and Eurasian plate is named as 'zone 1' and 'zone 2' is the intersection between Indian plate and Arabian plate. The Demarcation of seismogenic zone 1 using plate boundary and past seismicity is shown as example in the Fig. 3. The width of the seismogenic zones is different for each seismogenic zone based on the distribution of past events near the plate boundary. The minimum width is kept as 120 km on either side of the plate boundary. It can be observed from the Fig. 3 that the clusters of epicentres of past events very near to the plate boundary are considered as margin for seismogenic zones. The tectonic plates with very small area and the tectonic plates with very low seismic activity in the past are merged with the nearby seismogenic zones. It can be observed from the Figs. 2, 4 that the zones with very low seismic activity and intersections of very small tectonic plates are ignored. For example, the intersections between Sunda plate and Yangtze plate, Yangtze plate and Amur plate, Nazca plate and Cocos plate are eliminated. In the similar fashion, a total of 41 seismogenic zones are identified all over the globe and shown in the Fig. 4. All the 41 seismogenic zones are listed in the Table 1 along with their parenting plates. The events occurred inside these seismogenic zones are separated from the events occurred in
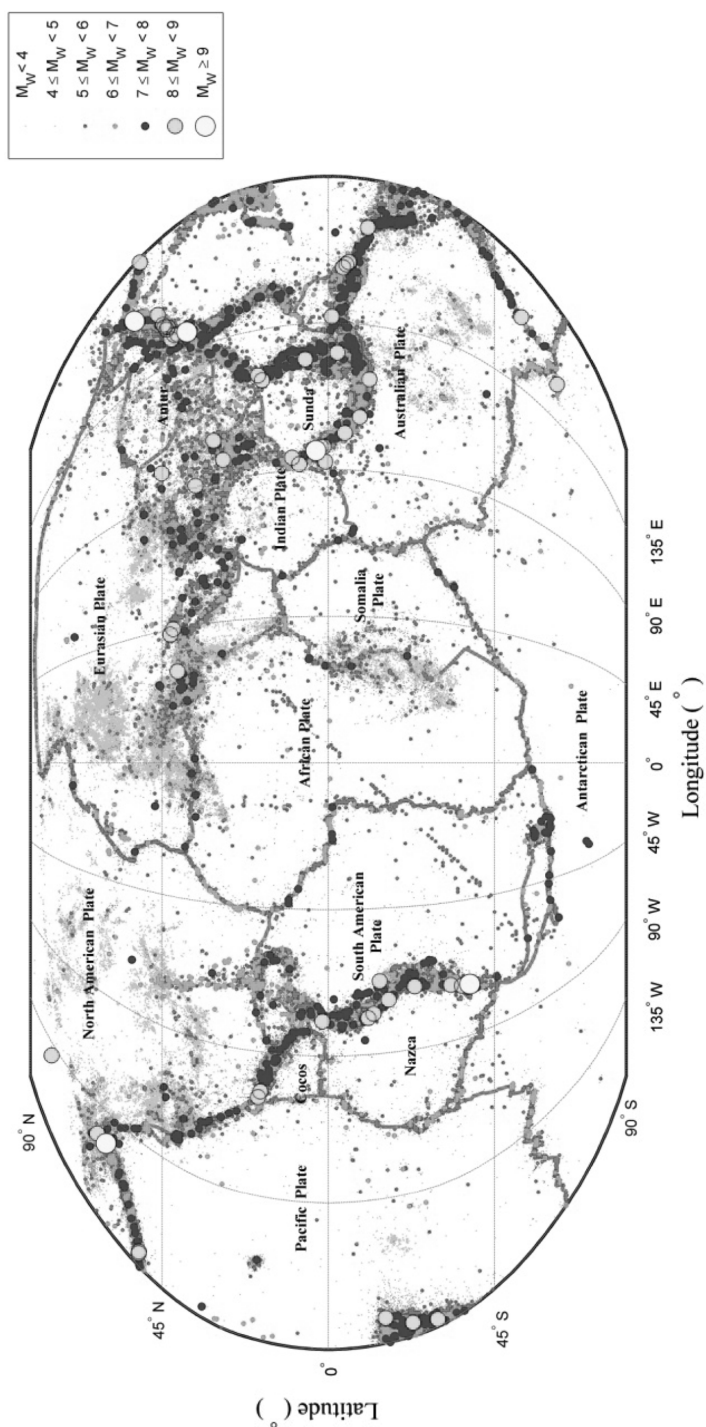
**Fig. 2** Epicentre distribution of earthquakes on tectonic plate boundaries
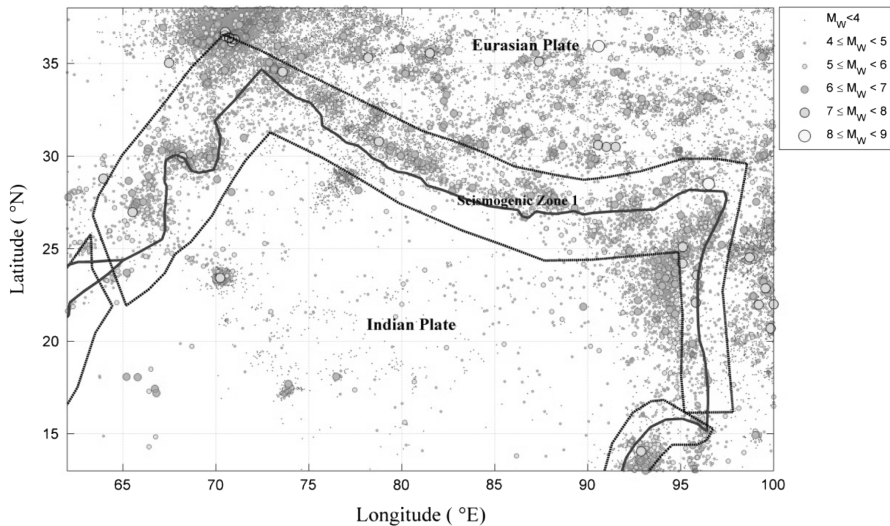
**Fig. 3** Demarcation of seismogenic zone 1 using plate boundary and past seismicity

the interior of tectonic plates. All over the globe, 66 % (3,220,870 earthquakes) of events are occurred in these 41 seismogenic zones. Whereas remaining 34 % (1,659,814 earthquakes) of events are occurred in other parts of the world. The zone 4 (intersection between Sunda plate and Philippine plate), zone 12 (Nazca plate-South American plate), zone 37 (North American plate-pacific plate), zone 38 (Okhotsk plate-pacific plate), zone 39 (Australian plate-Sunda plate) are the very active seismogenic zones. These seismogenic zones have more number of earthquakes as an indication of high seismic activity. Most of the deep earthquakes (>300 km) are accumulated only on these zones. From the past seismicity it is clear that, these regions are capable of producing large and deep earthquakes. The seismic activity is expected to be continuous in these 41 seismogenic zones. Hence it is necessary to study the seismicity pattern of these seismogenic zones.

## 3 Seismic energy time series

The magnitudes of individual events in the final catalogue (Fig. 1) are isolated and it is not physically meaningful to sum up for constructing a continuous time series. Hence it is convenient to convert the magnitudes into seismic energy and then sum up to create a continuous time series. The earthquakes occurred inside a particular seismogenic zone are singled out from the final catalogue (Fig. 1) and thus a separate catalogue is prepared for each seismogenic zone. The number of events in each of the seismogenic zone is listed in Table 1. The earthquake catalogue of India–Eurasian collision zone (zone 1) is shown in Fig. 5a. It can be observed from the figure that the catalogue is not homogeneous for all magnitudes. To obtain reliable results, the minimum magnitude of completeness ($M_c$) defined as the lowest magnitude above which 100 % of the events in a given region are detected has to be estimated. The magnitude of completeness ($M_c$) completeness of catalogue of each seismogenic zone is estimated by the procedure suggested by Wiemer and Wyss (2000). The maximum curvature method is used to determine $M_c$ from the frequency
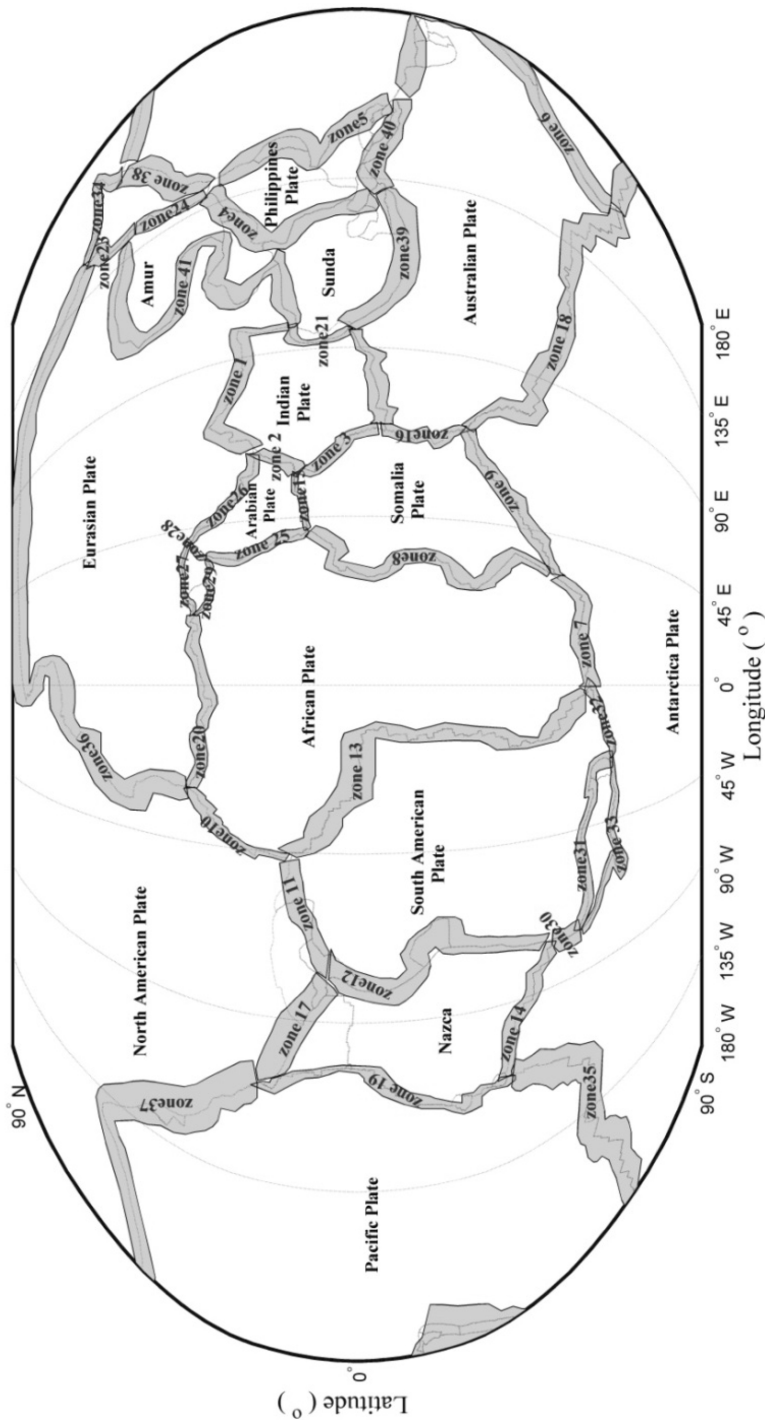
**Fig. 4** 41 Identified seismogenic zones on the plate boundaries

**Table 1** List of seismogenic zones and the performance of the modeling and forecasting strategy for annual seismic energy time series of seismogenic zones

| Zone number | Zone name | No. of earthquakes | Mc | Modeling period (1955–2004) | | | Forecasting period (2005–2014) | | | Expected seismic energy in 2015 (J) |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $\sigma_m$ | $CC_m$ | $PP_m$ | $\sigma_f$ | $CC_f$ | $PP_f$ | |
| 1 | Indian–Eurasian | 11,957 | 4.5 | 0.72 | 0.83 | 0.69 | 0.97 | 0.81 | 0.64 | $1.33 \times 10^{14}$ |
| 2 | Arabian–Indian | 134 | 5.0 | 0.61 | 0.87 | 0.76 | 0.84 | 0.90 | 0.77 | $4.62 \times 10^{9}$ |
| 3 | Somalian–Indian | 2065 | 5.0 | 0.73 | 0.88 | 0.77 | 0.78 | 0.82 | 0.55 | $1.06 \times 10^{11}$ |
| 4 | Sunda–Phillippines | 710,828 | 5.0 | 0.49 | 0.88 | 0.77 | 0.88 | 0.73 | 0.53 | $8.58 \times 10^{14}$ |
| 5 | Phillippines–pacific | 54,953 | 4.5 | 0.68 | 0.84 | 0.71 | 0.77 | 0.80 | 0.64 | $1.17 \times 10^{16}$ |
| 6 | pacific–Australian | 176,392 | 6.0 | 0.81 | 0.90 | 0.77 | 0.94 | 0.79 | 0.63 | $3.96 \times 10^{15}$ |
| 7 | Antarctican-African | 1120 | 6.0 | 0.73 | 0.90 | 0.81 | 0.87 | 0.85 | 0.58 | $1.86 \times 10^{13}$ |
| 8 | Somalia–African | 29,445 | 4.0 | 0.61 | 0.92 | 0.84 | 0.99 | 0.78 | 0.55 | $5.74 \times 10^{13}$ |
| 9 | Somalia–Antarctica | 2376 | 6.0 | 0.66 | 0.94 | 0.88 | 0.50 | 0.88 | 0.75 | $2.67 \times 10^{13}$ |
| 10 | Africa–North America | 3328 | 4.0 | 0.68 | 0.90 | 0.80 | 0.73 | 0.64 | 0.38 | $3.78 \times 10^{13}$ |
| 11 | North America–South America | 961 | 3.5 | 0.89 | 0.87 | 0.72 | 0.98 | 0.70 | 0.49 | $6.16 \times 10^{11}$ |
| 12 | Nazca–South America | 101,424 | 4.0 | 0.98 | 0.83 | 0.68 | 0.97 | 0.59 | 0.40 | $3.22 \times 10^{16}$ |
| 13 | South America-Africa | 5957 | 5.0 | 0.55 | 0.96 | 0.92 | 0.53 | 0.80 | 0.64 | $2.06 \times 10^{14}$ |
| 14 | Nazca–Antarctica | 2079 | 5.5 | 0.45 | 0.96 | 0.93 | 0.72 | 0.74 | 0.54 | $6.49 \times 10^{13}$ |
| 15 | Somalian–Arabian | 6163 | 5.0 | 0.76 | 0.85 | 0.72 | 0.95 | 0.61 | 0.41 | $6.68 \times 10^{14}$ |
| 16 | Somalian–Astralian | 2420 | 5.5 | 0.66 | 0.91 | 0.82 | 0.97 | 0.65 | 0.45 | $1.79 \times 10^{13}$ |
| 17 | Nazca–North American | 74,200 | 3.0 | 0.93 | 0.87 | 0.75 | 0.78 | 0.90 | 0.78 | $4.57 \times 10^{14}$ |
| 18 | Australian–Antarctican | 5610 | 4.0 | 0.39 | 0.95 | 0.91 | 0.61 | 0.63 | 0.38 | $6.70 \times 10^{13}$ |
| 19 | Nazca-pacific | 4411 | 6.0 | 0.43 | 0.97 | 0.95 | 0.58 | 0.59 | 0.42 | $3.40 \times 10^{13}$ |
| 20 | African–Eurasian | 65,461 | 3.5 | 0.88 | 0.91 | 0.82 | 0.99 | 0.63 | 0.39 | $7.45 \times 10^{13}$ |
| 21 | Sunda–Indian | 4854 | 6.5 | 0.98 | 0.88 | 0.74 | 0.97 | 0.65 | 0.39 | $2.04 \times 10^{15}$ |
| 22 | Indian–Australian | 581 | 6.0 | 0.98 | 0.82 | 0.67 | 0.99 | 0.67 | 0.43 | $3.67 \times 10^{13}$ |
| 23 | Eurasian–Okhotsk | 1008 | 6.5 | 0.86 | 0.93 | 0.82 | 0.93 | 0.81 | 0.59 | $4.00 \times 10^{12}$ |
| 24 | Okhotsk-Amur | 132,386 | 5.5 | 0.84 | 0.93 | 0.86 | 0.91 | 0.66 | 0.41 | $1.99 \times 10^{14}$ |
| 25 | African-Arabian | 29,917 | 5.0 | 0.90 | 0.89 | 0.79 | 0.89 | 0.57 | 0.44 | $6.46 \times 10^{11}$ |
| 26 | Arabian–Eurasian | 15,656 | 4.0 | 0.63 | 0.89 | 0.78 | 0.92 | 0.92 | 0.72 | $9.82 \times 10^{16}$ |
| 27 | Eurasian–Anatolia | 102,000 | 4.5 | 0.92 | 0.90 | 0.78 | 0.93 | 0.74 | 0.53 | $1.30 \times 10^{15}$ |
| 28 | Anatolia–Arabian | 12,305 | 5.0 | 0.62 | 0.94 | 0.84 | 0.95 | 0.71 | 0.49 | $3.15 \times 10^{12}$ |
| 29 | African–Anatolia | 37,962 | 5.0 | 0.62 | 0.90 | 0.81 | 0.89 | 0.41 | 0.31 | $1.19 \times 10^{11}$ |
| 30 | Antarctican–South America1 | 122 | 5.0 | 0.58 | 0.90 | 0.81 | 0.99 | 0.82 | 0.64 | $1.84 \times 10^{11}$ |
| 31 | South America–Scotia | 6195 | 4.5 | 0.93 | 0.90 | 0.81 | 0.97 | 0.45 | 0.35 | $9.89 \times 10^{13}$ |
| 32 | Antarctican–South America2 | 421 | 5.0 | 0.89 | 0.90 | 0.81 | 0.93 | 0.56 | 0.36 | $3.90 \times 10^{13}$ |

**Table 1** continued

| Zone number | Zone name | No. of earthquakes | Mc | Modeling period (1955–2004) | | | Forecasting period (2005–2014) | | | Expected seismic energy in 2015 (J) |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $\sigma_m$ | $CC_m$ | $PP_m$ | $\sigma_f$ | $CC_f$ | $PP_f$ | |
| 33 | Scotia–Antarctican | 1519 | 4.5 | 0.84 | 0.91 | 0.82 | 0.96 | 0.52 | 0.42 | $1.57 \times 10^{12}$ |
| 34 | Okhotsk-North American | 2946 | 4.0 | 0.78 | 0.90 | 0.81 | 0.85 | 0.76 | 0.53 | $2.84 \times 10^{11}$ |
| 35 | pacific-Antarctican | 2457 | 5.0 | 0.58 | 0.95 | 0.90 | 0.56 | 0.63 | 0.39 | $1.52 \times 10^{14}$ |
| 36 | Eurasian–North American | 17,123 | 4.0 | 0.47 | 0.91 | 0.83 | 0.64 | 0.81 | 0.64 | $8.32 \times 10^{13}$ |
| 37 | North American–pacific | 1,115,346 | 2.5 | 0.79 | 0.91 | 0.80 | 0.39 | 0.93 | 0.84 | $7.28 \times 10^{14}$ |
| 38 | Okhotsk–pacific | 365,890 | 5.0 | 0.70 | 0.92 | 0.84 | 0.89 | 0.80 | 0.64 | $1.73 \times 10^{14}$ |
| 39 | Sunda–Australian | 61,016 | 6.0 | 0.73 | 0.86 | 0.73 | 0.92 | 0.76 | 0.55 | $1.37 \times 10^{16}$ |
| 40 | Australian–Phillippines | 31,306 | 6.0 | 0.80 | 0.79 | 0.62 | 0.78 | 0.85 | 0.69 | $5.10 \times 10^{15}$ |
| 41 | Eurasian–Amur | 18,576 | 4.0 | 0.63 | 0.95 | 0.90 | 0.96 | 0.57 | 0.29 | $4.24 \times 10^{13}$ |

*Mc* agnitude completeness, $\sigma_m$ root mean square error in modelling period, $CC_m$ correlation coefficient for modelling period, $PP_m$ performance parameter for modelling period, $\sigma_f$ root mean square error in forecasting period, $CC_f$ correlation coefficient for forecasting period, $PP_f$ performance parameter for forecasting period

magnitude distribution. The magnitude of completeness is taken as the magnitude when the negative slope trend of the data stabilizes to approximate a straight line. The $M_c$ value for the seismogenic zone 1 is obtained as $M_w$ 4.5, as shown in Fig. 5b. The obtained magnitude completeness ($M_c$) of all the 41 seismogenic zones are reported in the Table 1. The events above $M_c$ in all the 41 catalogues are converted from the moment magnitude $M_W$ to seismic Moment $M_o$ using the empirical relation (Hanks and Kanamori 1979).

$$log_{10}(M_o) = \frac{3}{2}(M_w + 6.0). \tag{1}$$

Then, the seismic moments are converted into seismic energy ($E_S$) using the empirical equation derived by Choy and Boatwright (1995).

$$E_S = 1.6 \times 10^{-5} M_o. \tag{2}$$

The seismic energy releases of individual events in a particular year are added and the yearly seismic energy time series is constructed from 1950 to 2014 for all the 41 seismogenic zones. In order to model the small oscillations in the time-series, the $\log(E_S)$ time series which has no sharp peaks is used in the present article. The annual seismic energy time series for the Indian–Eurasian collision zone is shown in Fig. 5c. From the Fig. 5c, the peaks in energy time series can be related to the large earthquakes occurred in those specific years. The peak in the year 1950 corresponds to Assam earthquake of magnitude $M_W$ 8.6. The time series got a peak in the year 2005 because of the Kashmir earthquake of magnitude $M_W$ 7.5. The peak in the year 2013 corresponds to Pakistan earthquake of magnitude Mw 7.6. The average annual seismic energy release is calculated $3.79 \times 10^{15}$ J with the standard deviation of $2.80 \times 10^{16}$ J. The non-stationary nature of the annual earthquake energy time series can be observed from the Fig. 5c. Skewness and kurtosis are
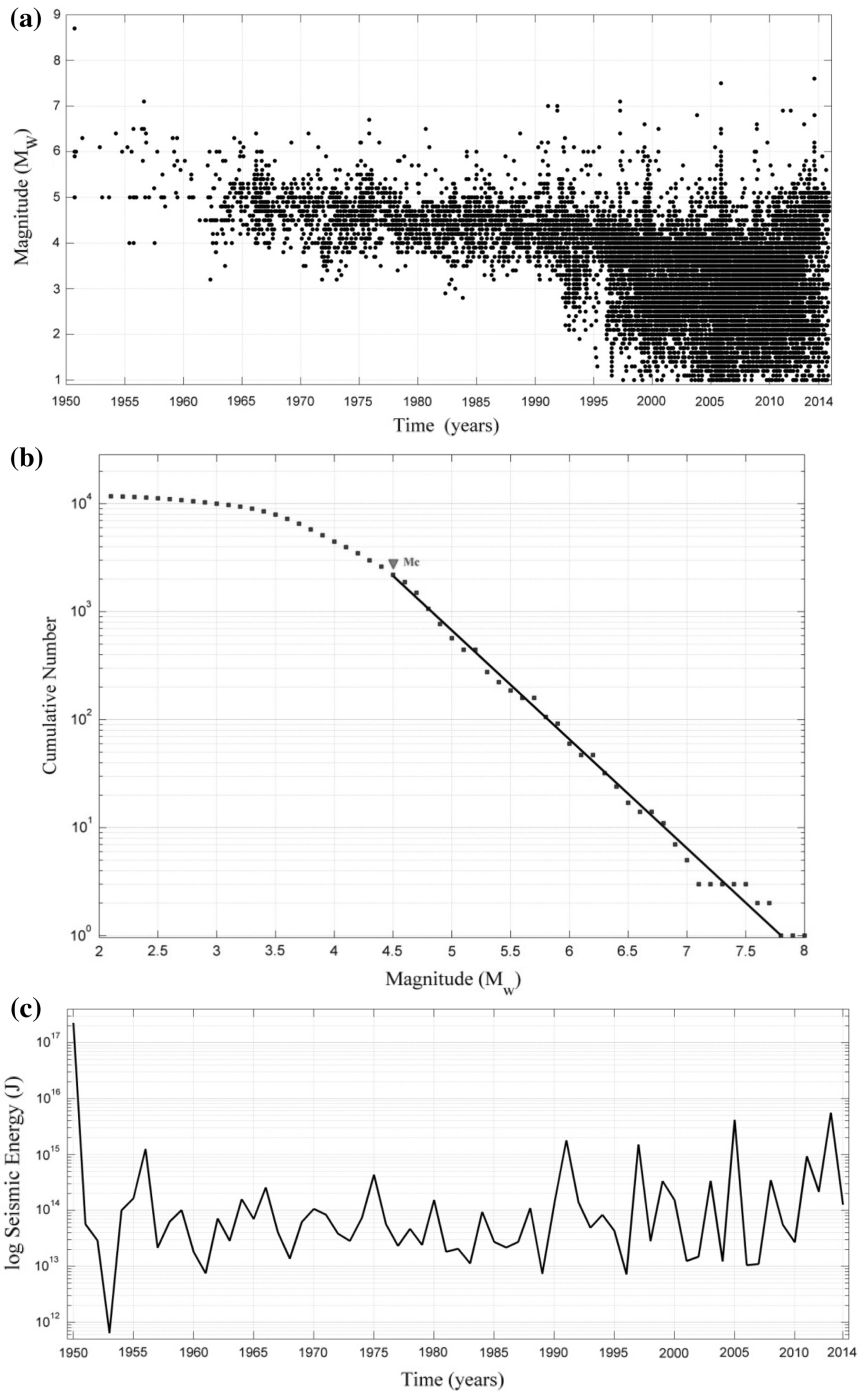
Fig. 5 **a** Catalogue of seismogenic zone 1. **b** Estimation of magnitude completeness for seismogenic zone 1. 5 **c** Annual seismic energy time series of seismogenic zone 1

higher-order statistical characteristics of a time series. The skewness is an indicator of the symmetry in the probability density function of the amplitude of time series. It measures deviation from a Gaussian distribution. The skewness will be zero, if the time series has equal number of small and large amplitudes. The skewness of the earthquake energy time series is estimated as 7.9 which indicate that the process is Non-Gaussian. The positive skewness indicates that data is right-skewed which is because of the 1950 Assam earthquake of magnitude $M_W$ 8.6. Kurtosis is a measure of whether the time series is peaked or flat compare to a normal distribution. The Kurtosis of the energy time series is 62.89. The time series with high kurtosis have a distinct peak near the mean and have thicker tails. There is high probability for extreme values. Hence Kurtosis value also confirms the Non-Gaussian nature of the time series.

# 4 Ensemble empirical mode decomposition

The annual seismic energy time series of seismogenic zones are appear to be non-linear and non-stationary (Fig. 5c). The non-stationary nature of the seismic energy time series is confirmed by many authors in the past (Liritzis and Tsapanos 1993). Therefore it is more scientific to use the Hilbert Huang Transform (HHT) proposed by Huang et al. (1998) which is capable of analyzing non-stationary data. The HHT is more adaptive to data since it does not use any predetermined function forms. The HHT consists of empirical modal decomposition (EMD) and Hilbert Transform. The EMD is a decomposition algorithm used to decompose the data into basis functions called IMFs. The extracted IMFs are adaptive, complete and orthogonal functions. Wu and Huang (2009) developed ensemble empirical mode decomposition (EEMD) technique to circumvent the mode mixing problem in EMD technique. In the EEMD technique, a white noise of finite amplitude is added to the data then EMD technique is used to extract the IMFs. An IMF is a narrow band time series with number of extremas and number of zero crossings differs by at most one. It has slowly varying frequency and amplitude. The mean value of the envelopes defined by the local maxima and minima of an IMF is zero. The IMFs are extracted from the data using a sifting process which consists of an interpolation method for constructing envelopes and a stopping criterion. The algorithm for extracting IMFs from the data using EEMD technique is as follows: First, white noise of finite amplitude $[w_h(t)]$ is added to the data $E_S(t)$ which lead to the new data $E(t)$. Then the consecutive maxima and minima in $E(t)$ are identified to construct the lower and upper envelopes by cubic splines. The average of the positive and negative envelopes $h(t)$ is determined at every time step. This average which is the local mean of the data is subtracted from the data to get $E_1(t) = E(t)–h_0(t)$. This new data $E_1(t)$ is treated as in the previous step $E_2(t) = E_1(t)–h_1(t)$. This process is repeated n times till the sieved data $E_n(t)$ becomes an IMF (i.e. $IMF_1$). To extract the $IMF_2$, the $IMF_1$ is subtracted from the data $E(t)$ and the whole shifting process is repeated. In a similar way, $IMF_3$, $IMF_4$… are extracted until the sieved data shows no oscillations. After extracting all the IMFs the entire EMD procedure is repeated $M$ times by adding a different white noise time series to the data, where $M$ is called as an ensemble number. An ensemble of IMFs will be generated through this process and the ensemble mean of these IMFs will lead to the final empirical modes. In the present article, two important parameters, the ratio of the standard deviation of the white noise to that of data and the ensemble number have to be fixed before the sifting process. The ratio of the standard deviation of added noise to that of data should not be too low to introduce enough changes in the extremas of the decomposed

data. In contrast, very high ratio will lead to physically insignificant IMFs. Similarly, very high ensemble number needs higher computation cost whereas smaller number will not be enough for the cancellation of noise in IMFs. In the present article, the ratio is taken as 2 and ensemble number is fixed as 5000 by trial and error in order to maintain consistency. A total of 6 IMFs are obtained from each time series using the fixed parameters.

In Fig. 6, all the six IMFs along with the energy time series of seismogenic zone 1 are shown as sample. The IMFs are shown in the order in which they are extracted. It can be noted from the Fig. 6 that the IMFs are having equal number of extremas and zero crossings and its envelopes are also symmetric with respect to zero line. The last IMF ($IMF_6$) indicates the trend of the earthquake energy time series of seismogenic zone 1. It is clear from the Fig. 6 that the energy release in the year 1950 (Assam earthquake $M_W$ 8.6) controls the shape of all the IMFs. The sum of all the six IMFs will lead to the annual earthquake energy time series of seismogenic zone 1.

The derived IMFs of all seismogenic zones are having similar properties of sine or cosine waves. Hence it is easy to estimate their time period by counting the number of extrema in an IMF. In this fashion, the periods of all IMFs of 41 seismogenic zones are calculated and their minimum and maximum ranges are given in Table 2 along with the mean values. It can be observed from the Table 2 that the estimated periods of seismic energy time series of seismogenic zones are in the range of 2.5–3, 4.5–6, 8.5–12, 14–20, 17–22 and 21–32 years. Liritzis and Tsapanos (1993) have also reported the dominant periods of global earthquake energy time series as 3(±0.5), 4.5, 6.5, 8–9, 14–20 and 31–34 years. It can be noted that their periods are almost same as the periods observed in the present study. The periods of IMFs of few seismogenic zones are listed in Table 3. Another important statistical parameter called percentage variance, which is the ratio of variance of each IMF to the data variance is also estimated for all seismogenic zones and included in Tables 2 and 3. The percentage of variance of IMF indicates the contribution of each IMF to annual earthquake energy release of particular zone. The spatial distribution of the periods of first four IMF's and their percentage variance are reported in Figs. 7, 8, 9 and 10. It can be observed from the Fig. 7a that period of first IMF varies from 2.53 to 3 with an average value of 2.75 years. The seismogenic zone 29 has the maximum period of $IMF_1$ (3.01 years) whereas zone 13 shows the minimum period of $IMF_1$ (2.53 years). The $IMF_2$ has periods from 4.83 to 5.93 with a mean value of 5.28 years (Fig. 8a). The minimum period of $IMF_2$ (4.83 years) is observed for the seismogenic zones 22 and 38. The maximum period of $IMF_2$ is estimated as 5.93 years for the seismogenic zone 3. The periodicity of $IMF_3$ of 41 seismogenic zones oscillating from 8.55 years to 10.89 years (Fig. 9a). The minimum period of $IMF_3$ is estimated for seismogenic zone 19 and the maximum value is estimated for the zone 38. The time period of $IMF_4$ varies from 13.83 to 20.37 years respectively (Fig. 10a). The average time period of $IMF_4$ of all seismogenic zones is estimated as 18.03 years. The minimum time period is observed for zone 8 whereas the seismogenic zone 9 has maximum time period of $IMF_4$. It is clear from the Table 2; Figs. 7, 8, 9 and 10 that the $IMF_1$ is the predominant mode with an average period of 2.75 years is contributing more than 50 % (Fig. 7b) to the annual seismic energy for all the seismogenic zones. The $IMF_1$ of the seismogenic zone 24 provides major contribution (76.02 %) towards annual seismic energy release. The $IMF_1$ of the zone 19 has low contribution (33.76 %) to the seismic energy release. The $IMF_2$ is the second important mode which has around 17 % (Fig. 8b) contribution with an average period of 5 years. The $IMF_2$ of seismogenic zone 22 and 12 have minimum and maximum contributions respectively when compared to other seismogenic zones. The $IMF_3$ is contributing about 6–20 % (Fig. 9b) and its period oscillates around 11 years which is same as 11 years
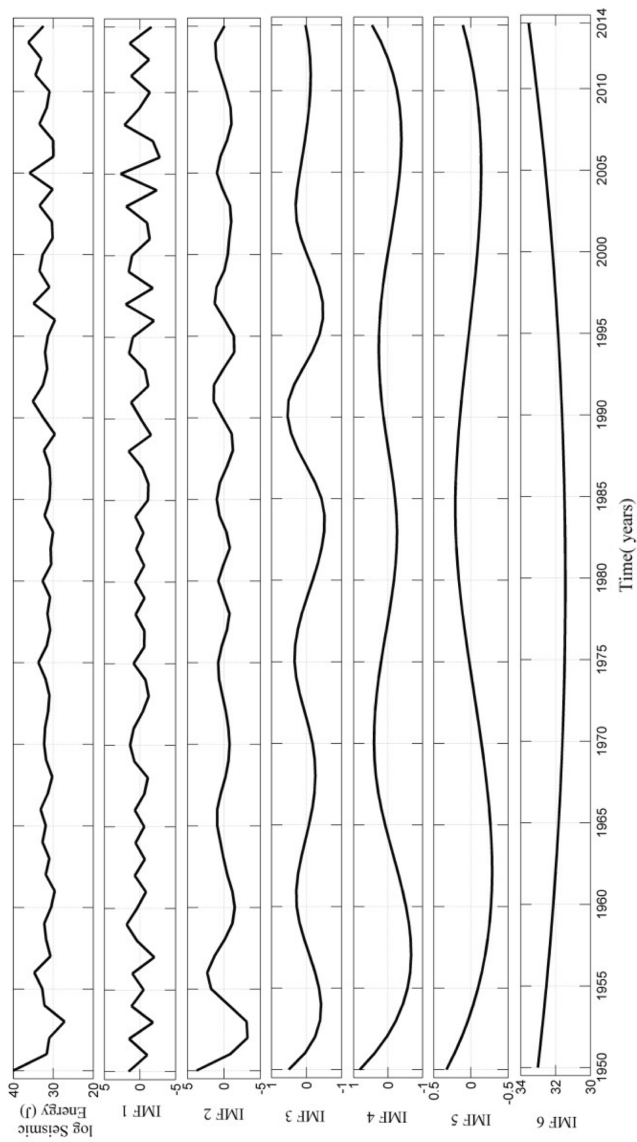
**Fig. 6** IMFs of yearly seismic energy time series of zone 1

**Table 2** Range of period and percentage variances of IMFs of seismic energy time series of 41 seismogenic zones

| S.no | IMFs | Period (years) | | | Percentage variance (%) | | |
|---|---|---|---|---|---|---|---|
| | | Min. | Mean | Max. | Min. | Mean | Max. |
| 1 | $IMF_1$ | 2.53 | 2.75 | 3.01 | 33.76 | 52.47 | 76.02 |
| 2 | $IMF_2$ | 4.83 | 5.28 | 5.93 | 7.22 | 16.91 | 26.94 |
| 3 | $IMF_3$ | 8.55 | 10.81 | 10.89 | 5.85 | 10.23 | 19.51 |
| 4 | $IMF_4$ | 13.83 | 18.03 | 20.37 | 3.18 | 8.14 | 10.55 |
| 5 | $IMF_5$ | 17.12 | 20.85 | 22.13 | 2.81 | 5.17 | 8.71 |
| 6 | $IMF_6$ | 21.33 | 28.09 | 32 | 2.42 | 10.28 | 23.79 |

**Table 3** Periods and percentage variances of IMFs observed for seismic energy time series of few seismogenic zones

| IMFs | Zone 1 (Indian–Eurasian) | | Zone 3 (Somalian–Indian) | | Zone 37 (North American–pacific) | |
|---|---|---|---|---|---|---|
| | Period (years) | Per. vari. (%) | Period (years) | Per. vari. (%) | Period (years) | Per. vari.(%) |
| $IMF_1$ | 2.62 | 44.30 | 2.72 | 49.31 | 2.81 | 45.11 |
| $IMF_2$ | 5.72 | 16.13 | 5.93 | 15.14 | 5.41 | 17.50 |
| $IMF_3$ | 10.45 | 12.48 | 10.67 | 10.52 | 10.86 | 12.41 |
| $IMF_4$ | 17.40 | 7.62 | 18.50 | 8.74 | 17.15 | 9.25 |
| $IMF_5$ | 22.13 | 5.27 | 20.75 | 4.61 | 21.44 | 4.71 |
| $IMF_6$ | 32.00 | 16.59 | 32.00 | 13.04 | 32.00 | 12.66 |

period of sunspot cycle (http://sidc.oma.be/sunspot-data/). The $IMF_3$ of seismogenic zone 20 has very low contribution (5.8 %). The $IMF_3$ of zone 12 contributed around 20 % which is higher than other seismogenic zones. The $IMF_4$ has contribution in the range of 3–11 % (Fig. 10b) with an average time period of 18 years. This period can be related with the Lunar standstill cycle of 18.6 years. The $IMF_4$ of seismogenic zone 16 has very less contribution (3 %) whereas zone 32 has maximum percentage variance (11 %). The $IMF_5$ is contributing around 5 % with an average period of 21 years. The time period and percentage variance of IMFs remain constant while considering the whole globe as a single zone or separate seismogenic zones (Kavitha and Raghukanth 2015). This consistency emphasizes the influence of sun spot and lunar standstill cycles on earthquake occurrences in these seismogenic zones. The IMFs of the earthquake energy time series are simple and well-behaved compare to the data. Hence it would be interesting to forecast the IMFs instead of the complex time series. But 41 separate models have to be developed to forecast the seismic energy release of seismogenic zones. The correlated zones can be grouped together and the single forecasting model can be used for all zones in a group. This grouping can be done with the help of PCA technique and it is explained in the next section.

# 5 Regionalization of seismogenic zones using PCA

The seismogenic zones correlated among themselves can be grouped together which is known as regionalization. The time series data of seismogenic zones of same group, instead of individual zone data, could be used to improve the forecasting. Moreover, it
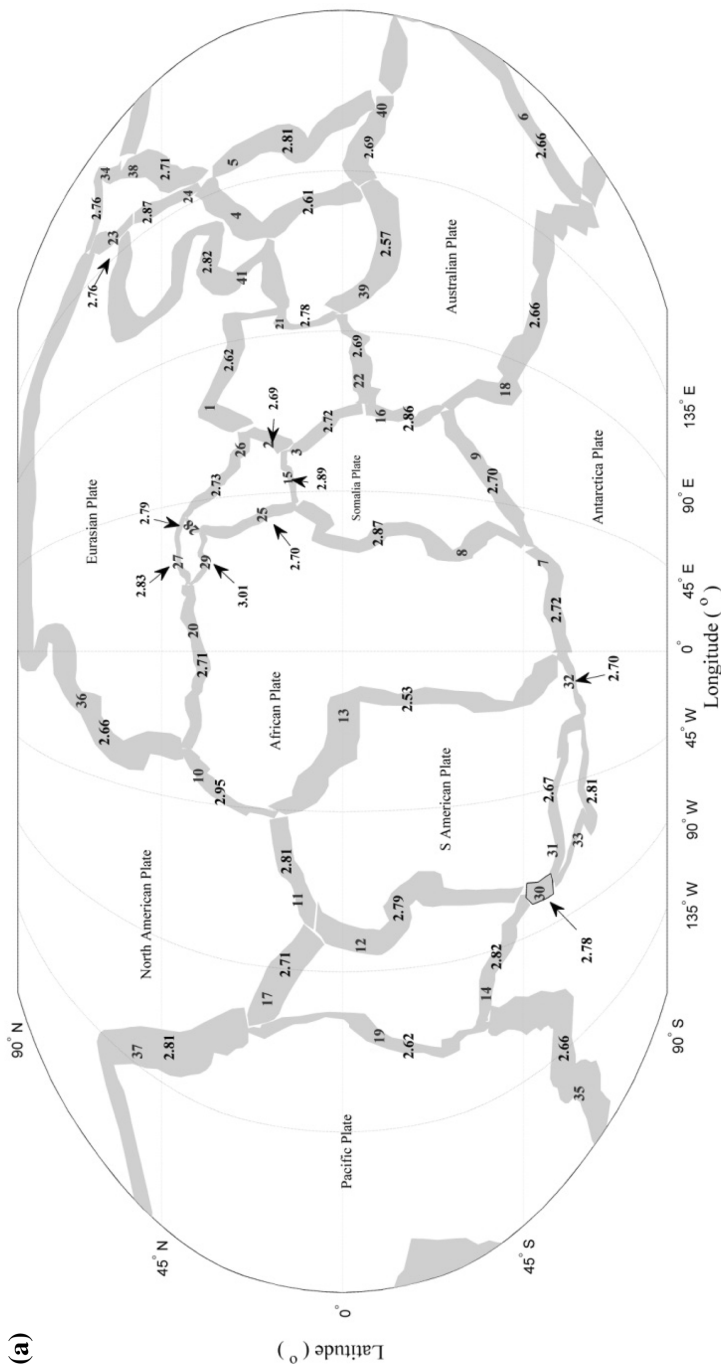
**Fig. 7** **a** Spatial distribution of periods of IMF$_1$ of the annual seismic energy time series of seismogenic zones. **b** Spatial distribution of percentage variances of IMF$_1$ of the annual seismic energy time series of seismogenic zones
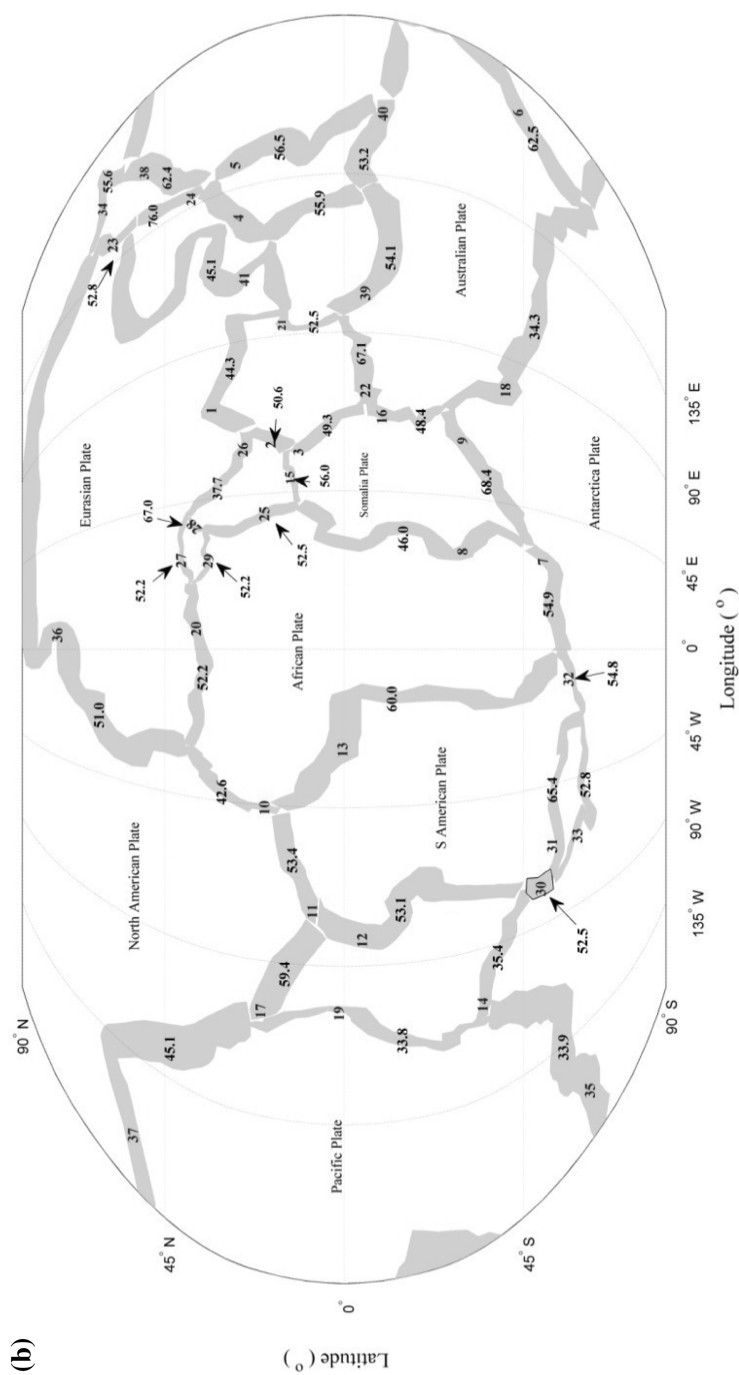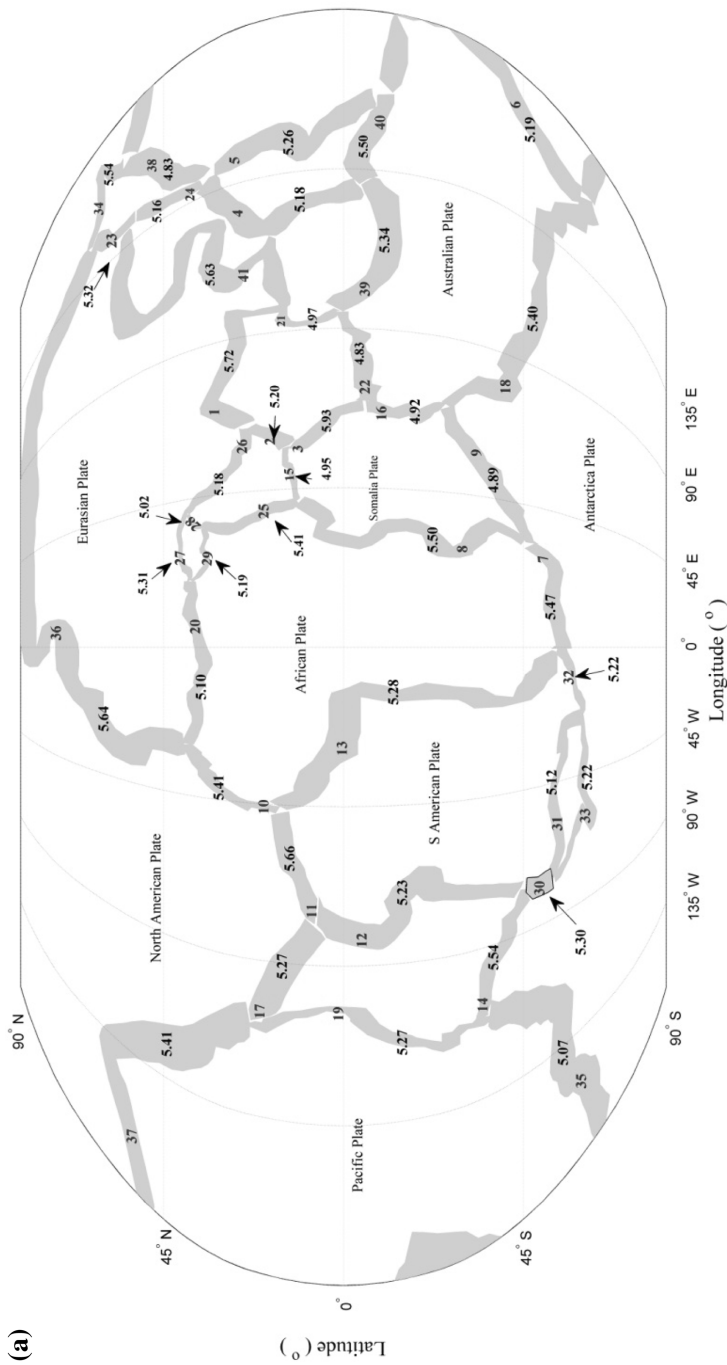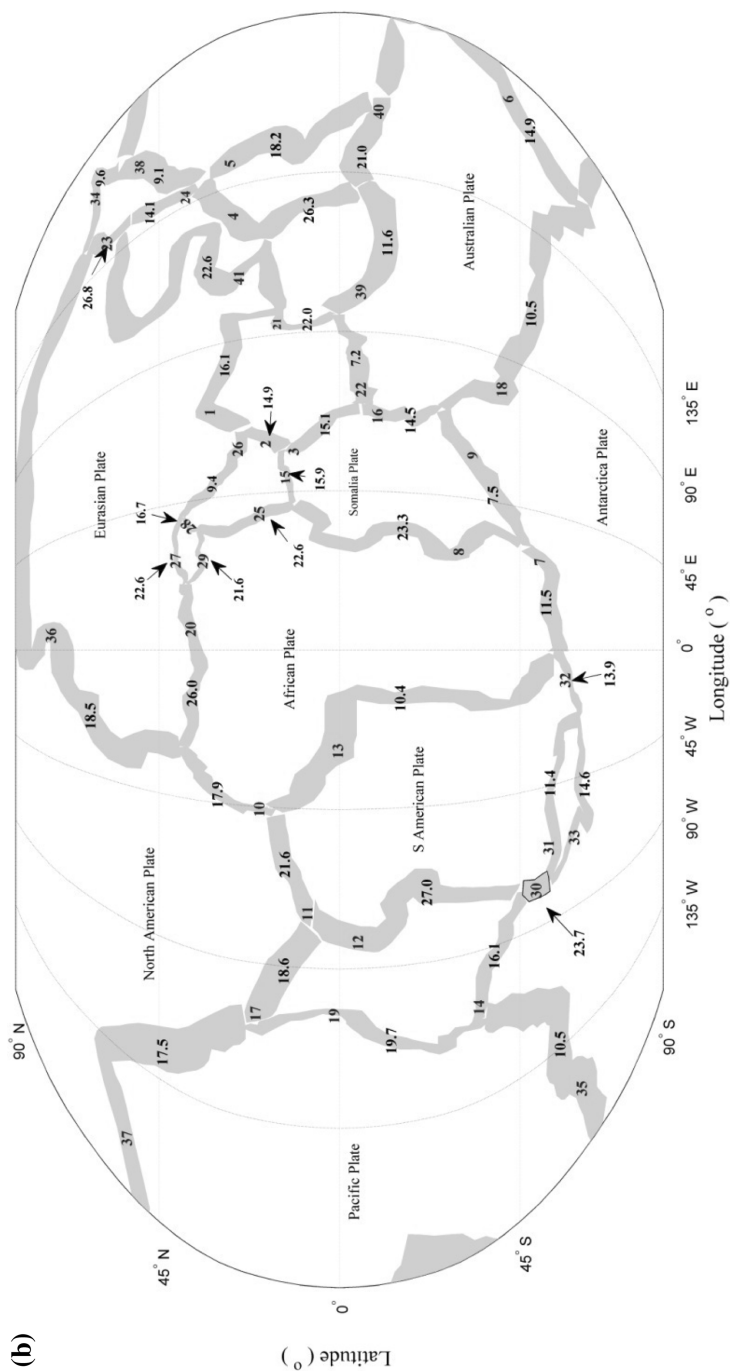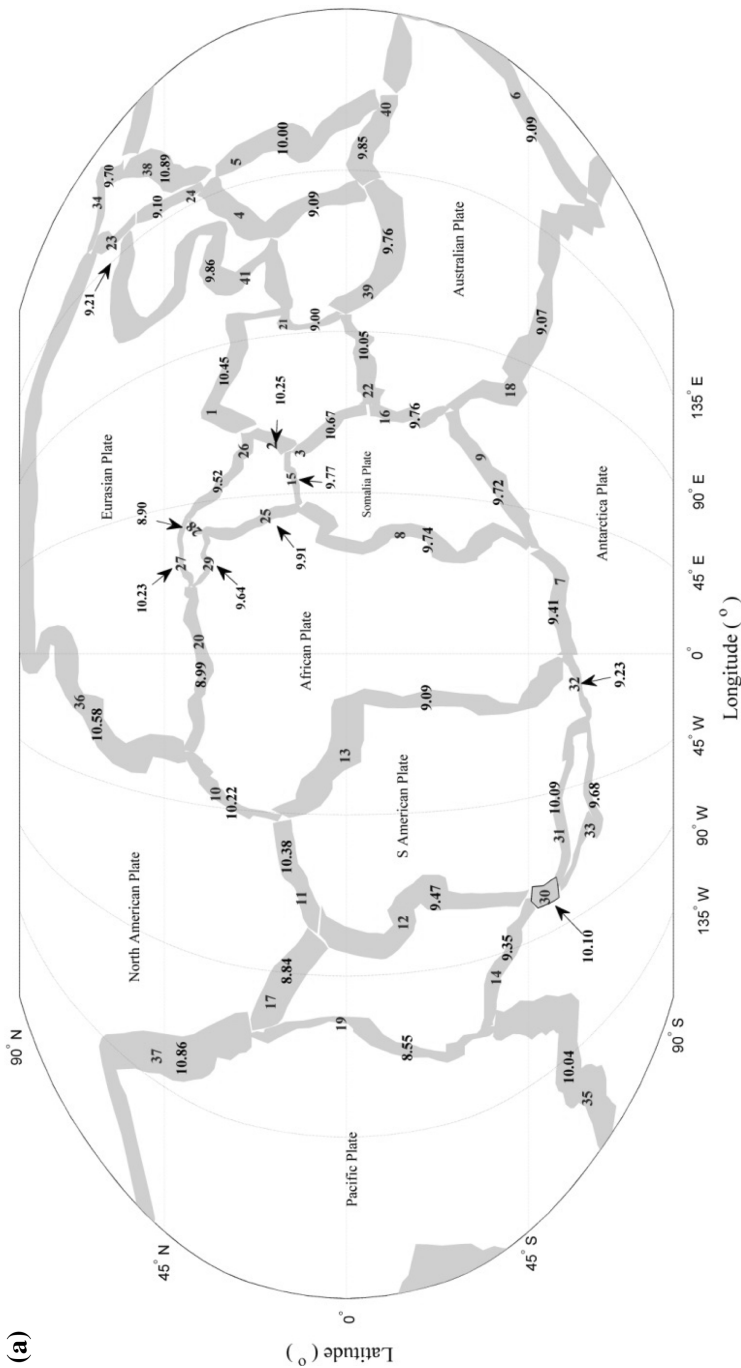
**Fig. 7** continued

**Fig. 8** **a** Spatial distribution of periods of $IMF_2$ of the annual seismic energy time series of seismogenic zones. **b** Spatial distribution of percentage variances of $IMF_2$ of the annual seismic energy time series of seismogenic zones
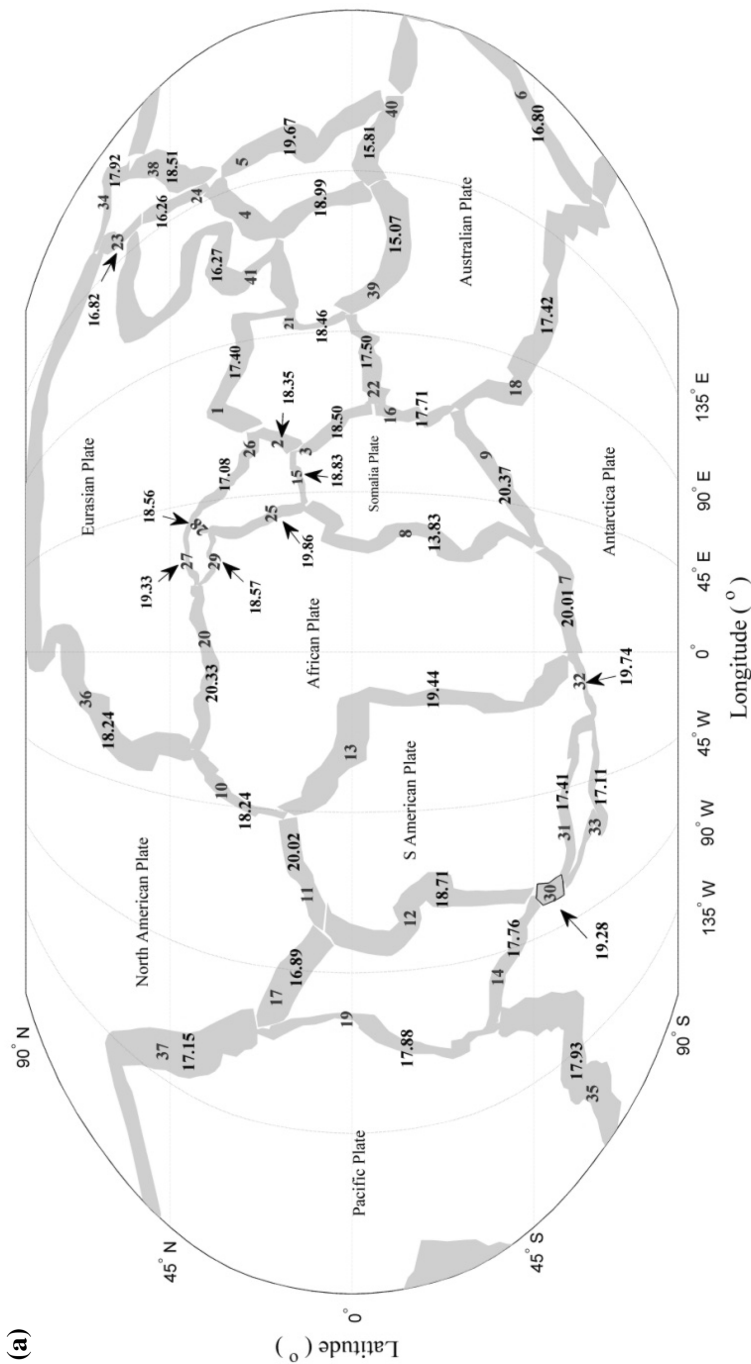
**(b)**

Fig. 8 continued

**Fig. 9** **a** Spatial distribution of periods of IMF$_3$ of the annual seismic energy time series of seismogenic zones. **b** Spatial distribution of percentage variances of IMF$_3$ of the annual seismic energy time series of seismogenic zones

**Fig. 9** continued

**Fig. 10** **a** Spatial distribution of periods of IMF$_4$ of the annual seismic energy time series of seismogenic zones. **b** Spatial distribution of percentage variances of IMF$_4$ of the annual seismic energy time series of seismogenic zones
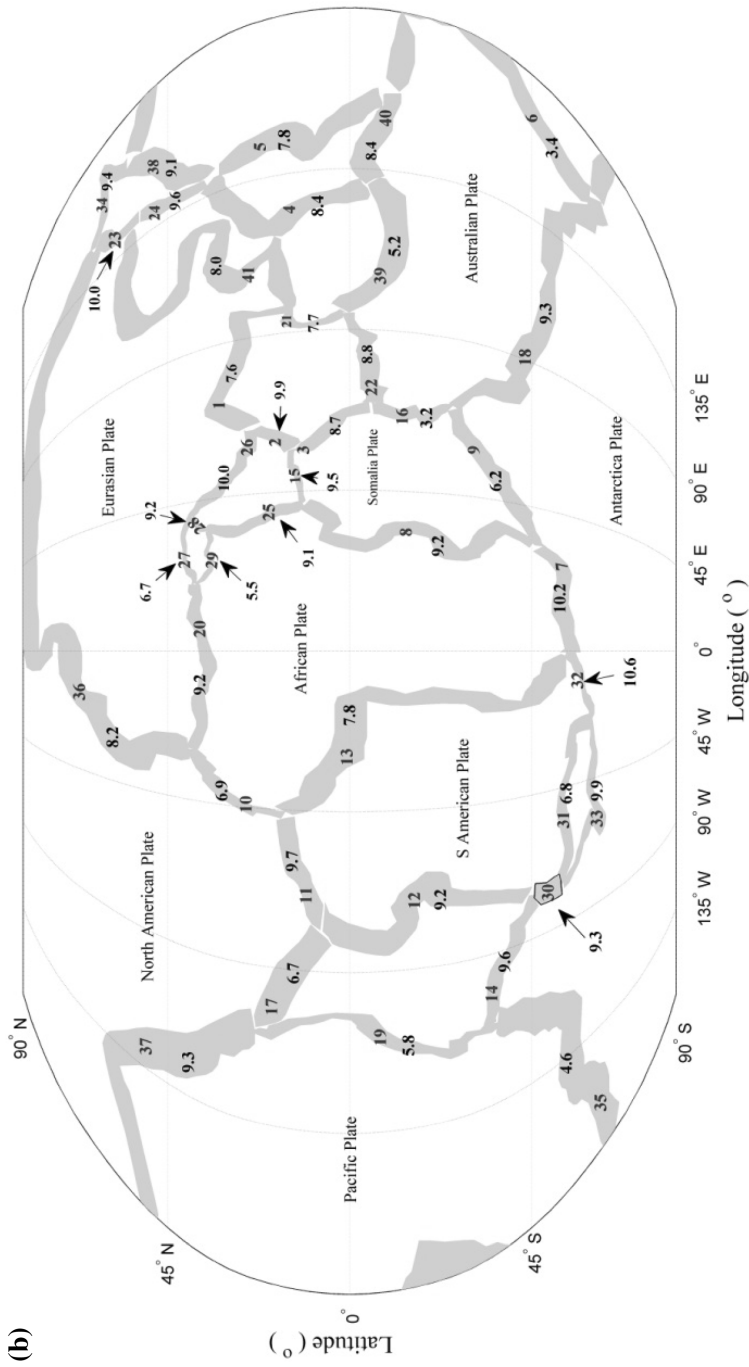
**Fig. 10** continued

enhances the signals in the data on larger spatial scales (Nicholson 1986). In the present article, the log(Es) time-series of 41 seismogenic zones are used for regionalization. If the seismic energy release at zone $i$ ($i = 1,2,...m$) in the year $t$ ($t = 1,2,...n$) is $S_{it}$, the covariance matrix of the data is

$$C_{ij} = \left(\frac{1}{N}\right) \sum_{t=1}^{n} s_{it}s_{jt}, \tag{3}$$

where $s_{it}$, the centred data of zone $i$, is

$$s_{it} = (S_{it} - q_i), \tag{4}$$

where

$$q_i = \left(\frac{1}{N}\right) \sum_{t=1}^{n} S_{it}. \tag{5}$$

The eigenvalues $\lambda_j$ and corresponding eigenvectors $\{\phi_{ij}\}$ of the symmetric matrix $C_{ij}$ are derived. The $j$th principal component for year $t$ is

$$P_{jt} = \sum_{i=1}^{n} s_{it}\phi_{ij} \ (j = 1, 2, \ldots m). \tag{6}$$

The eigenvalues are normalized as

$$\overline{\lambda_j} = \left(m\lambda_j\right) \Big/ \sum_{j=1}^{m} \lambda_j. \tag{7}$$

Here, Eq. (4) represents an orthogonal decomposition of the seismic energy release. Hence the principal components $p_{jt}$ should be carrying the temporal signatures that present in the actual data. It is important to decide how many of these PCs are significant. Then those can be related to the zonal seismic energy release data to arrive at homogenous groups. The amount of the variance of data explained by a principal component indicates its importance. The first 10 principal components are explained more than 90 % of the data variance, which are significant principal components. Figure 11 shows the time series of 10 significant principal components.

$$s_{it} = \sum_{j=1}^{m} P_{jt}\phi_{ij}. \tag{8}$$

The first ten $p_{jt}$ series contain the maximum temporal information of $s_{it}$ data. One can do regionalization depending on how the zone data series and the above $p_{jt}$ series are related among themselves. The correlation between time series of all 41 seismogenic zones and the $p_{jt}$ are estimated. All the zones that are maximally correlated with the same sign with respect to the same principal component are grouped together. Dyer (1975) used similar approach for identifying homogeneous rainfall regions in South Africa. In the present study, a significance level is set up as a criterion for the correlation between the principal components and the seismogenic zones. The PCs are derived from the annual seismic energy series of seismogenic zones. Since, these two are expected to be correlated. If any two variables with sample size $n$ to be deducted as correlated, then the minimum level of significance should be $\rho = \pm 2/\sqrt{n}$. Here sample size $n = 65$ and the minimum
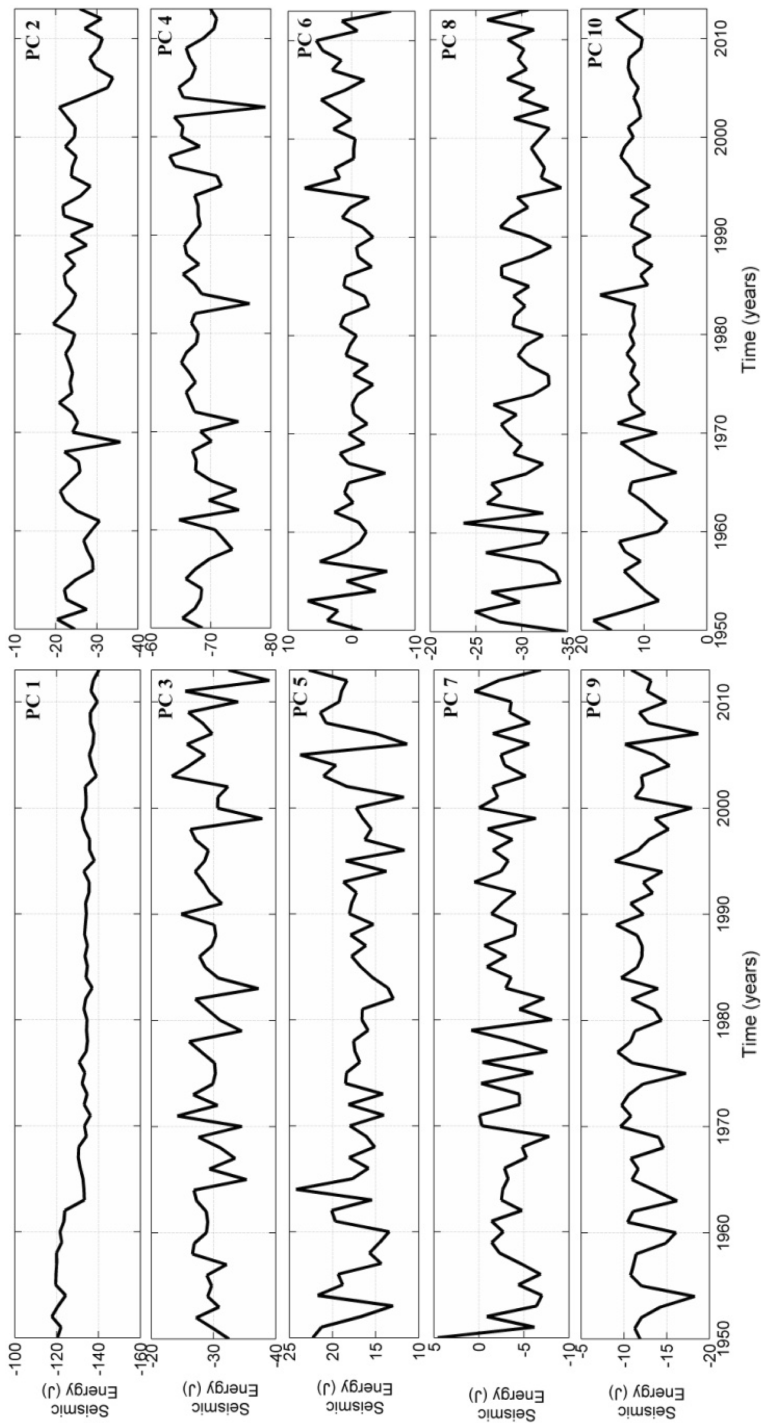
**Fig. 11** The significant principal component time series

| Table 4 Membership details of the principal groups | Group | No. of zones | Seismogenic zone numbers |
|---|---|---|---|
| | 1 | 15 | 3, 7, 9, 13, 14, 15, 16, 18, 19, 24, 31, 32, 33, 35, 41 |
| | 2 | 3 | 8, 11, 21 |
| | 3 | 2 | 22, 27 |
| | 4 | 2 | 10, 23 |
| | 5 | 1 | 34 |
| | 6 | 3 | 1, 17, 36 |
| | 7 | 3 | 12, 26, 40 |
| | 8 | 2 | 6, 25 |
| | 9 | 1 | 20 |
| | 10 | 1 | 39 |
| | 11 | 1 | 38 |
| | 12 | 2 | 29, 30 |
| | 13 | 2 | 5, 28 |
| | 14 | 1 | 4 |
| | 15 | 1 | 2 |
| | 16 | 1 | 37 |

significance limit is $|\rho| \geq 0.25$ (Hays and Winkler 1970). The zones satisfying this criterion are selected and grouped together. When a zone is found to be correlated with more than one PC, it is assigned to the PC with higher correlation value. Thus all the 41 seismogenic zones are grouped under 16 principal groups. The membership details of corresponding principal groups are listed in Table 4. It can be noted from the Table 4 that the first principal group is the biggest group containing 15 seismogenic zones. The second, sixth and seventh principal groups have three seismogenic zones as members. The first seismogenic zone (Indian–Eurasian intersection) is in the sixth principal group. The principal groups 5, 9, 10, 11, 14, 15 and 16 are having single seismogenic zone as member. The grouped seismogenic zones are shown in the Fig. 12. The seismogenic zones belong to the same group are colored alike to exhibit the grouping. It is clear from the Fig. 12 that the demarcated principal groups are not geographically contiguous. But most of the seismogenic zones below equator are belong to the first principal group. Popoola and Hammed (2007) have studied global seismicity pattern by dividing the whole globe into zones of width 10° with the equator as the centre. They reported that all the zones below equator are having similar trend with increased seismicity. Their observations are in agreement with the present study, where most of the zones below equator are correlated among themselves and belong to the same group. As the correlation among seismic energy release of seismogenic zones is clearly understood, forecasting of the same can be attempted.

# 6 Forecasting

Various strategies are proposed in the literature for the prediction of earthquakes. But the forecast of the seismic energy release can directly point out the maximum possible magnitude of future event. The annual earthquake energy time series of any seismogenic zone is supposed to have the influences of all the causes within the time series itself. The
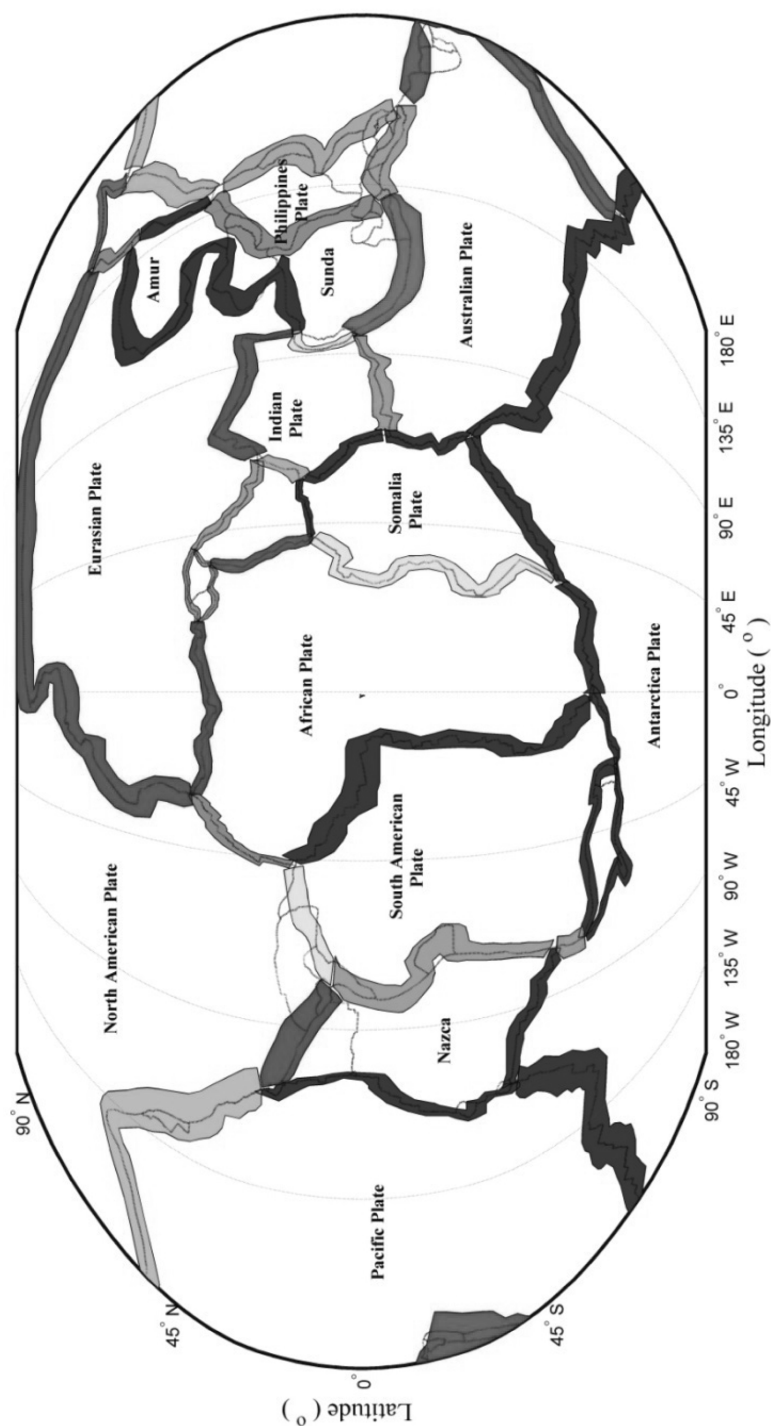
**Fig. 12** Grouped seismogenic zones

energy time series can be modeled as a function of past values with the help of sufficiently long data. Hence an autoregressive model which is capable of reproducing the past can be used to forecast the energy time series a year ahead. As the time series is highly complex, it is easy to model and forecast the IMFs which are simpler than the data. The forecasted values of the IMFs can be added to estimate an expected energy release. It is difficult to find the $IMF_1$ at the end of the data because the envelope on the both sides of the end point is not defined. To overcome this difficulty, the previous value of the end point can be used as the next value. But this method is not convincing for forecasting problem. If the data $S_t$ is available then IMFs can be extracted only for t = 2, 3, 4 …. n − 1. With the increase in the distance between each extrema, the extrapolation errors will spread into the signal and misrepresent the higher IMFs at the end points. Iyengar and Raghukanth (2005) have modeled the linear and non-linear parts separately to circumvent this end problem of IMFs. In the present article, this difficulty has overcome by modeling the $IMF_1$ and the remaining part of the data i.e. $Y_t = (S_t\text{-}IMF_{1t})$ separately. As both the parts of the data are non-linear, they can be effectively modeled using ANN technique (Wong et al. 2000). Eisner and Tsanis (1992) have proved that ANN works for modeling and extending the chaotic trajectories of the Lorenz equation. The ANN has three layers (input, hidden and output layers) of simple processing units interconnected by acyclic links. Selecting the number of hidden layers, number of nodes in the hidden layer and type of activation function play a significant role in model construction.

The seismogenic zones with correlation among themselves are arranged into single group (Table 4 and Fig. 12). Thus the time series data of all the seismogenic zones in a particular group can be used for forecasting instead of individual zone data. The correlation among the time series of zones will enhance the signal in the data and improve the predictability of the model. Hence, instead of 41 separate ANNs, 16 ANNs for each group is sufficient to forecast all the 41 seismogenic zones. For the present study, after many trials, an ANN model with a single hidden layer with five nodes as shown in Fig. 13a is chosen for modeling the first part of the data. The input layer has five nodes which depend on the past four values of $Y_t$, an end value of the data $S_t$. For the second part of the data that is $Z_t = S_t$-modeled $(Y_t)$, the same type of ANN model which depends on the past five values of the $Z_t$ is selected as shown in Fig. 13b. The mathematical representation of the relationship between input parameters $(S_t, Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4})$ and output parameter $(Y_t)$ of first part is as follows,

$$Y_t = bias_{OL}^{I} + \sum_{l=1}^{5} w_l^{I} \cdot g\left(bias_{HL,l}^{I} + w_{1,l}^{I}S_t + \sum_{k=2}^{5} w_{k,l}^{I}Y_{t-k}\right). \tag{9}$$
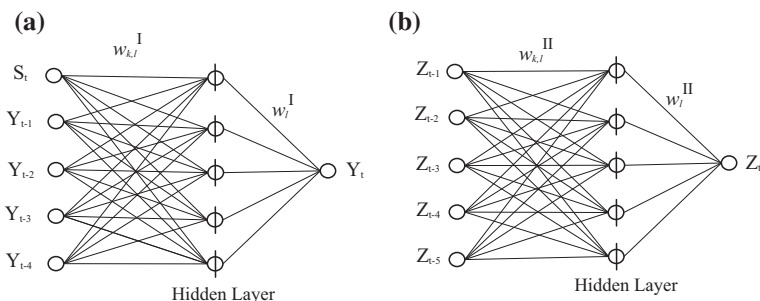


**Fig. 13  a** ANN architecture for modeling part I $(Y_t)$. **b** ANN architecture for modeling part II $(Z_t)$

Similarly, the relationship between input parameters ($Z_{t-1}$, $Z_{t-2}$, $Z_{t-3}$, $Z_{t-4}$, $Z_{t-5}$) and output parameter ($Z_t$) of second part is as follows,

$$Z_t = bias_{OL}^{II} + \sum_{l=1}^{5} w_l^{II} \cdot g\left(bias_{HL,l}^{II} + \sum_{k=1}^{5} w_{k,l}^{II} Z_{t-k}\right), \qquad (10)$$

where $bias_{OL}^{I}$ and $bias_{OL}^{II}$ are output layer bias values of part I and part II; $w_l^{I}$ and $w_l^{II}$ are weight factor between neuron $l$ of the hidden layer and single output neuron for part I and II; $g(\cdot)$ is a hyperbolic tangent sigmoid transfer function adopted between the input layer and hidden layer; $bias_{HL,l}{}^{I}$ and $bias_{HL,l}{}^{II}$ are bias value at neuron $l$ of the hidden layer in ANN of part I and part II; $w_{k,l}^{I}$ and $w_{k,l}^{I}$ are weight factors between the input variables ($k = 1, 2, ..., 5$) and neuron $l$ of the hidden layer in ANN of part I and part II. A linear function is adapted between hidden and output layer. Thus, this network is equivalent to a nonlinear autoregressive model. Equations 9 and 10 indicate one output node in the output layer, which is used for one-step-ahead forecasting. The Levenberg-Marquardt back propagation training algorithm is implemented for training the model. The 36 unknowns in Eqs. 9 and 10 are estimated by minimizing the mean square error between the actual data and the values simulated by the model. The weight factors and bias values of the best-fitted network for all the seismogenic zones in Group 6 is given as a sample in Tables 5, 6, 7 and 8. The group 6 contains three seismogenic zones that are zone1 (intersection between Indian plate and Eurasian plate), zone 17 and zone 36. The weight matrix and bias values between the input layer and hidden layer of ANN model for part I and part II of Group 6 are given in Tables 5, 7 respectively. Similarly, the weight factors and bias value between the hidden layer and output layer of ANN model for part I are provided in Table 6 and the same for part II are given in Table 8. These values can be used for modeling annual seismic energy time series any one of the seismogenic zones in the group 6 using Eqs. 9 and 10.

To demonstrate the ability of the proposed model, the first 50 years (1955–2004) of data is taken as modeling period and the last 10 (2005–2014) years of data is taken as testing period. Three statistical parameters are selected for verifying the performance of the model. First, the RMS error $\sigma_m(\varepsilon)$ is calculated for the model. Second, the $CC_m$ between the original data and the proposed model is checked. Then, the $PP_m = 1 - (\sigma_m^2/\sigma_d^2)$ is also verified, where $\sigma_m^2$ is the mean squared error and $\sigma_d^2$ is the variance of actual data. These three statistical parameters of both ANN model for all the 41 seismogenic zones are included in Table 1. For an ideal model, $\sigma_m(\varepsilon)$ should be zero and $CC_m$ and $PP_m$ should be

**Table 5** Weight factors and bias values between the input layer and hidden layer in ANN for modeling part I of Group 6

| Weight factors | Number of hidden neurons ($l$) | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| $w_{1,l}^{I}$ | 11.0961 | −14.8741 | −36.9565 | −3.0563 | 3.2496 |
| $w_{2,l}^{I}$ | −101.9641 | 11.6871 | −33.2123 | 2.7305 | 8.9608 |
| $w_{3,l}^{I}$ | −21.3657 | −5.6051 | 29.0972 | −1.2096 | 3.8783 |
| $w_{4,l}^{I}$ | 0.6175 | −82.6819 | −37.6910 | −10.6771 | −15.9095 |
| $w_{5,l}^{I}$ | −37.9652 | −123.3606 | −80.6621 | 58.4872 | −28.3047 |
| $bias_{HL}^{I}$ | −6.8350 | −58.6693 | −65.2335 | 22.8773 | −7.1276 |

**Table 6** Weight factors and bias value between the hidden layer and output layer in ANN for modeling part I of Group 6

| Weight factor | Number of hidden neurons ($l$) | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | $Bias_{OL}^{I}$ |
| $w_l^{I}$ | 6.7562 | −0.3057 | 10.2119 | −0.1722 | 49.7032 | −28.3258 |

**Table 7** Weight factors and bias values between the input layer and hidden layer in ANN for modeling part II of Group 6

| Weight factor | Number of hidden neurons ($l$) | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| $w_{1,l}^{II}$ | 3.3427 | −27.3785 | 39.5646 | −0.7887 | 14.4044 |
| $w_{2,l}^{II}$ | 9.2116 | −26.6067 | 35.7449 | −0.4128 | −11.4428 |
| $w_{3,l}^{II}$ | 3.8865 | 12.1310 | 31.0907 | 1.1419 | 5.4977 |
| $w_{4,l}^{II}$ | −16.2800 | −141.2651 | 39.9519 | −0.0091 | 81.8434 |
| $w_{5,l}^{II}$ | −28.8714 | −239.2606 | 85.0636 | 6.4297 | 122.3780 |
| $bias_{HL}^{II}$ | −7.2642 | −121.6197 | 69.3621 | 2.7426 | 58.1102 |

**Table 8** Weight factors and bias value between the hidden layer and output layer in ANN for modeling part II of Group 6

| Weight factor | Number of hidden neurons ($l$) | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | $Bias_{OL}^{II}$ |
| $w_l^{II}$ | −20.3593 | −0.0806 | 2.0656 | 0.1114 | −0.0663 | −0.2954 |

unity. It can be noted from Table 1 that for most of the seismogenic zones $CC_m$ and $PP_m$ are close to unity and $\sigma_m(\varepsilon)$ also near to zero. It indicates the satisfactory performance of the proposed model. The statistical parameters $\sigma_f(\varepsilon)$, $CC_f$ and $PP_f$ are calculated for forecasting period i.e. last 10 (2005–2014) years and included in the Table 1. To be significant, the $CC_f$ should be greater than 0.5. The predictability of the proposed model is beyond the significant level for most of the seismogenic zones.

In the present article, future seismic energy release of a particular seismogenic zone in a group is considered to depend on the past seismic energy releases of all other seismogenic zones in that group. This association enhanced some of the signals present in data, which leads to the improved performance of the model. The excellent performance of the model can be observed from the Table 1, particularly from the zone 2, zone 9, zone 17, zone 26 and zone 37. But some the seismogenic zones like zone 29, zone 31 and zone 41 have very low predictability which can be attributed to the complexities involved in the process and short modeling period. The largest expected seismic energy release in the forecasting period is estimated as $4.74 \times 10^{17}$ J for the seismogenic zone 39 (intersection between Sunda and Australian plate) in the year 2006. However, the actual seismic energy released from the seismogenic zone 39 in the year 2006 is $7.46 \times 10^{15}$ J. One can compare the expected and actual magnitude of large events in the forecasting period. In the year 2005,

Sumatra earthquake (epicentre 2.07°N, 97.02°E) occurred in seismogenic zone 39. The expected seismic energy release of the seismogenic zone 39 is estimated as $3.78 \times 10^{16}$ J with the standard deviation of $8.85 \times 10^{17}$ J. The actual magnitude of Sumatra earthquake is $M_w$ 9.2 whereas the expected magnitude is 8.2. Similarly, 2011 Tohoku earthquake (epicentre 38.32°N, 142.37°E) occurred in seismogenic zone 38. The expected seismic energy release for the seismogenic zone 38 is estimated as $2.26 \times 10^{16}$ J with the standard deviation of $6.32 \times 10^{17}$ J and the expected magnitude is Mw 8.1. However, the actual magnitude of Tohoku earthquake is Mw 9.0. It can be noted that although the model could not predict exact magnitude, the actual magnitudes are within the known error band. In the Fig. 14, the modeled and actual energy releases are compared year by year for both training period and forecasting period to demonstrate the skill of the model. The training period has the CC of 0.97 which is much higher than significant value of 0.28. The proposed forecasting strategy is also highly efficient with the CC of 0.84 which is higher than the significant value of 0.63.

In Fig. 15, the comparison between actual data and predicted earthquake energy time series of seismogenic zones 1 is shown. The model predicted the exact energy release in the years 2009, 2010 and 2011. Although there are considerable differences in the years 2005, 2007, 2008 and 2013, the forecasted values are within a known error band. As the future value depends on the past five values, the time series can be extended by 1 year. Hence it is possible to find the expected energy release in the year 2015 by extending the modeled time series. The forecasted value of energy time series is a random variable with model error as its standard deviation. For the year 2015, the energy release is estimated as $1.33 \times 10^{14}$ J with the standard deviation of $8.17 \times 10^{14}$ J. The annual seismic energy
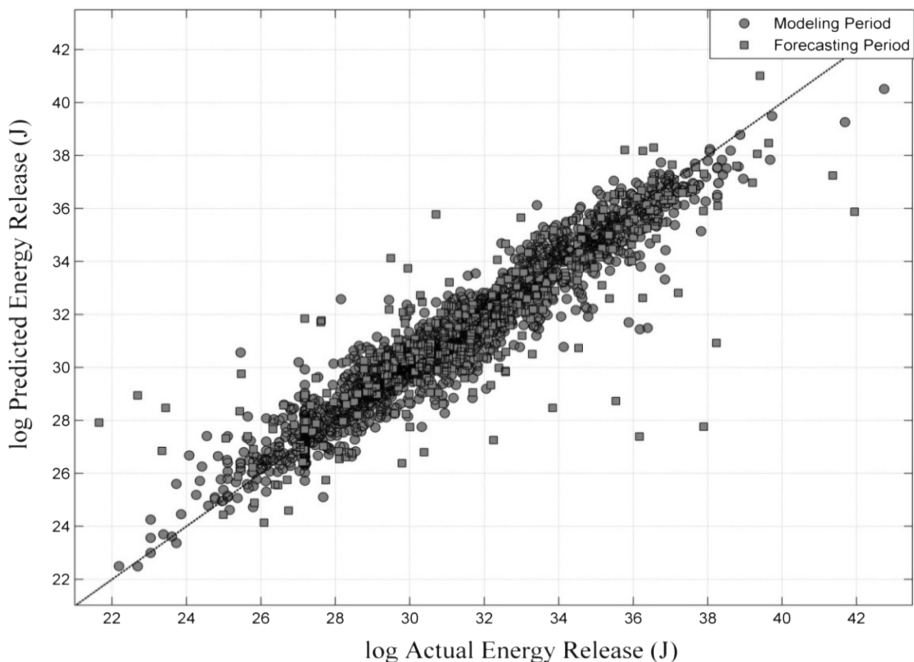


**Fig. 14** Comparison between actual energy releases and modeled energy releases during the training (1955 to 2004) and forecasting periods (2005–2014); (*dashed line* expected line)
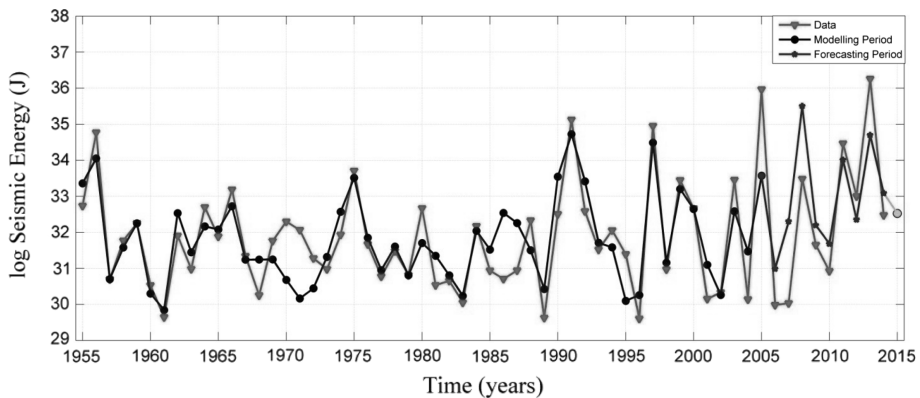
**Fig. 15** Comparison between actual data and prediction of earthquake energy time series of seismogenic zone 1

release will be greater than $1.6 \times 10^{16}$ J if there is a single event with magnitude $M_W \geq 8$ (Eqs. 1 and 2). Hence there is no possibility of occurrence for the events with magnitude $M_W \geq 8$ in the seismogenic zone 1 in year 2015. In the similar fashion, the expected energy release for the 2015 of all 41seismogenic zones are estimated and listed in the Table 1. The seismogenic zone 2 is expected to have minimum seismic energy release ($4.62 \times 10^9$ J) in the year 2015. It can be noted from the Table 1 that, the seismogenic zones 12 and 26 have a possibility of occurrence for the event with magnitude $M_W \geq 8$ in the year 2015. The seismogenic zone 26 (the intersection between Arabian and Eurasian tectonic plate) has largest expected seismic energy release in the year 2015. The seismic energy is estimated as $9.82 \times 10^{16}$ J with the standard deviation of $3.25 \times 10^{17}$ J for the zone 26.

## 7 Summary and conclusions

This article presents a methodology to understand the seismicity pattern of seismogenic zones and to forecast the earthquake occurrence of the seismogenic zones with the help of PCA technique. The ISC–GEM Global Instrumental Reference Earthquake Catalogue (http://www.isc.ac.uk/iscgem/index.php) has been utilized in the present article. A total of 41 active seismogenic zones are identified from tectonic plate boundaries using past seismicity. The annual earthquake energy time series is constructed by adding the energy releases of all the events in a particular year for each seismogenic zone. The EEMD technique is used to extract six IMFs from the earthquake energy time series. The periodicities and percentage variances of IMFs are estimated for each seismogenic zone. The correlation of third and fourth IMFs with solar and lunar cycles indicates the influence of these celestial bodies on earthquake energy release of seismogenic zones. The principal components are estimated from the annual seismic energy series of 41 seismogenic zones. The derived PCs are used to organize 41 seismogenic zones into 16 groups. Due to the complexities involved, the auto regressive models are used for modeling and forecasting the seismic energy release of seismogenic zones. After several trials, an ANN model with a single hidden layer and five input nodes has been selected for modeling the data. The time series data of all seismogenic zones of a group, instead of individual zone data, is used for

forecasting the future seismic energy release. This association leads to the improved performance of the model. The seismic energy release data from 1955 to 2004 has been used for modeling purpose and the data from 2005 to 2014 has been used for validation purpose. It is found that the developed model is efficient in forecasting the annual earthquake energy release of most of the seismogenic zones with a known error band. The expected energy release of all seismogenic zones in the year 2015 is estimated by extending the time series for 1 year. The estimated energy budget will be helpful for the government agencies and policy makers to know in advance about the earthquake occurrence with probable magnitude. It is found that the seismogenic zones 22 and 26 have a possibility of occurrence for events with magnitude $M_W \geq 8$ in the year 2015. But one cannot state anything regarding the occurrence of events with magnitudes $M_W$ 6 or 7 from this seismic energy budget. In the proposed model, the annual earthquake energy time series is used. One can attempt monthly or weekly seismic energy time series as a short term prediction. The magnitude completeness of the catalogue should be ensured for better predictability of the model. Instead of taking 50 years as modeling period, whole span of data can be used to train the model for better performance. In the present study, the model parameters are held constant throughout the forecasting period which can be updated every year to enhance the performance. It is important to compile the seismicity data to a greater extent because long span of data will improve the accuracy of the prediction.

# References

Bird P (2003) An updated digital model of plate boundaries. Geochem Geophys Geosyst 4(3):1–52

Bolt B (2005) Earthquakes: 2006 centennial update: the 1906 big one, Freeman. W. H. & Company, New York

Choy GL, Boatwright JL (1995) Global patterns of radiated seismic energy and apparent stress. J Geophys Res 100(B9):18205–18228

Dyer TGJ (1975) The assignment of rainfall stations into homogeneous groups: an application of principal component analysis. Q J R Meteorol Soc 101:1005–1013

Ehrendorfer M (1987) A regionalization of Austria's precipitation climate using principal components analysis'. J Climatol 7:71–89

Eisner JB, Tsonis AA (1992) Non-linear prediction, chaos, and noise. Bull Am Meteor Soc 73:49–60

Hanks TC, Kanamori H (1979) Moment magnitude scale. J Geophys Res 84(B5):2348–2350

Hays WH, Winkler RL (1970) Statistics: Probability, inference and decision. Rinehart and Winston, New York

Huang NE et al (1998) The empirical mode decomposition and the hilbert spectrum for nonlinear and non stationary time series analysis. Proc R Soc Lond A454(903):995

Iyengar RN, Basak P (1994) Regionalization of Indian monsoon rainfall and long-term variability signals. Int J Climatol 14:1095–1114

Iyengar RN, Raghukanth STG (2005) Intrinsic mode functions and strategy for forecasting Indian monsoon rainfall. Meteorol Atmos Phys 90:17–36

Kavitha R, Raghukanth STG (2015) Forecasting of global earthquake energy time series. Adv Adapt Data Anal (under review)

Kutzback JE (1967) Empirical eigenvectors of sea level pressure, surface temperature and precipitation complexes over North America. J Appl Meteorol 6:791–802

Liritzis I, Tsapanos TM (1993) Probable evidence for periodicities in global seismic energy release. Earth Moon Planet 60:93–108

Nicholson SE (1986) Spatial coherence of African rainfall anomalies: inter-hemispheric teleconnection. J Clim Appl Meteorol 25:1365–1379

Overland JE, Preisendorfo RW (1982) A significance test for principal components applied to cyclone climatology. Mon Wea Rev 110:1–4

Popoola OI, Hammed S (2007) Variation of global seismicity across the equator. Online J Earth Sci 1(3):136–138

Preisendorfer RW, Zwiers FW, Barnett TP (1981) Foundations of principal component selection rules, SIO Ref. Series 81–4, Scripps Institute of Oceanography, University of California

Srivastava HN, Bhattacharya SN (1998) 'Application of principal component analysis to some earthquake related data in the Koyna region, India. Eng Geol 50:141–151

Telesca L, Colangelo G, Hattori K, Lapenna V (2004) Principal component analysis of geoelectrical signals measured in the seismically active area of Basilicata Region (southern Italy). Nat Hazard Earth Syst Sci 4:663–667

Tsapanos TM (1998) Seismic hazard for some regions of the world examined using strain energy release. J Balk Geophys Soc 1(2):19–26

Tsapanos TM, Liritzis I (1992) Time-lag of the earthquake energy release between three seismic regions. Pure Appl Geophys 139(2):293–308

Varga P, Krumm F, Riguzzi F, Doglioni C, Süle B, Wang K, Panza GF (2011) Global pattern of earthquakes and seismic energy distributions: insights for the mechanisms of plate tectonics. Tectonophysics 530–531:80–86

Wiemer S, Wyss M (2000) Minimum magnitude of complete reporting in earthquake catalogs: examples from Alaska, the Western United States, and Japan. Bull Seism Soc Am 90:859–869

Wong BK, Vincent S, Jolie L (2000) A bibliography of neural network business applications research: 1994–1998. Comput Oper Res 27:1023–1044

Wu Z, Huang NE (2009) Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv Adapt Data Anal 1(1):1–41