Precursory correlation dimensions for three great earthquakes

P. N. S. Roy$^{1,*}$ and Sankar K. Nath$^2$

$^1$Department of Applied Geophysics, Indian School of Mines University, Dhanbad 826 004, India
$^2$Department of Geology and Geophysics, Indian Institute of Technology, Kharagpur 721 302, India

On 26 December 2004, the Great Sumatra earthquake of $M_c = 9$ rocked the Southeast Asian continent. Fractal analysis of earthquakes ($M_c \geq 4$) occurring during 1996–2005 led to the decrease of correlation dimension in the narrow time-span prior to the mainshock. Such precursors were also observed prior to the 17 August 1999 Izmit earthquake of $M_c = 7.8$ and the 3 November 2002 Denali earthquake of $M_c = 8.5$. Spatio-temporal clustering apparently indicates a highly stressed region, an asperity or weak zone from where the rupture propagation eventually nucleates causing a large earthquake. The results are in conformity with similar analysis for a few earthquakes reported earlier in India and elsewhere.

Keywords: Asperity, clustering, precursory correlation dimension, self-similarity.

The fractal spatial correlation dimension ($D_C$) provides a quantitative measure of the spatial clustering of events indicating the seismicity of a region$^{4,5}$. Experiments on microfracturing of rocks suggesting fractal nature of a series of 29 granite samples demonstrated that decrease in fractal dimension can be used as a predictor of rock failure$^6$.

In the present study, the correlation integral method$^6$ was used to determine correlation dimension. A ‘unique fingerprint’ of a multifractal object requires the introduction of an infinite hierarchy of fractal dimensions, known as generalized fractal dimensions as given below:

$$D_q = \frac{1}{(q-1)} \lim_{r \to 0} \left[ \frac{\log \left( \sum_i \{P_i(r)\}^q \right)}{\log r} \right],$$  \hspace{1cm} (1)

where $D_q$ exhibits a non-trivial scaling behaviour for different values of $q = 1, 2, 3, \ldots,$ and $P_i(r)$ is the probability that the events fall into a square box of length $r$.

Using eq. (1), one readily estimates the correlation fractal dimension for an integer $q = 2$ as a special case. The correlation dimension is therefore calculated as:

$$\text{Correlation dimension} = \lim_{q \to 2} D_q = D_2 = D_C.$$  \hspace{1cm} (2)

In two dimensions, $D_q$ approaching a value of 2 signifies a uniform coverage of the plane. Uniform distribution of events (objects) that are scale-invariant (self-similar) is termed as monofractal.

Temporal variations of $D_C$ were studied considering 50 events in consecutive windows for the Sumatra–Andaman region; Izmit, Turkey, and Denali Fault, Alaska region.

Method

In this article, the correlation integral method was utilized to determine spatial correlation dimension $D_C$ and its variation with time. For this purpose, 50 events in consecutive windows were considered.

The spherical triangle method has been used to determine the distance between the two epicentres$^{9,4,7}$. The correlation integral approach has been employed to compute the correlation dimension.

The fractal correlation dimension was derived from the correlation integral$^{8,6,9}$, which is a cumulative correlation function that measures the fraction of points in the two-dimensional space and is defined as,

$$C(r) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} H(r - r_{ij}),$$  \hspace{1cm} (3)

where $N$ is the total number of pairs of vectors with respect to one another in the fractal set to determine $D_C$ (for 50 events window, $N$ will be $^{50}C_2$, i.e. 1225), $r$ the length scale, $r_{ij}$ the distances between the points of a set which is obtained through spherical triangle method explained above and $H$ the Heaviside step function. Therefore, $C(r)$ is proportional to the number of pairs of points of the fractal set separated by a distance less than $r$. If the system of points examined is a fractal set, the graph of $C(r)$ in logarithmic coordinates must be a linear function with slope $D_C$ equal to the fractal dimension of the system. The graph of $C(r)$ at different stages of the fracture process is shown in Figure 1. The curves show a clear self-similar behaviour in a wide range of about two orders of magnitude on the space scale. Deviations from linear dependence in the range of large scales are connected with the finite size of samples, while the other deviation in the range of small scales reflects the boundary effect of data for the region of investigation.

*For correspondence. (e-mail: pns_may1@yahoo.com)
The $D_c$ value is inversely related to the degree of clustering and it requires higher degree of accuracy in both space and time of occurrence of events, as the present analysis depends on the spatio-temporal distribution of earthquake sequences. It is standard practice to consider the completeness magnitude as the threshold for any analysis that uses a well-constrained earthquake catalogue, prohibiting error due to the presence of low-magnitude earthquakes in the region. In the present analysis, we attempted the computation of $D_c$, that is dependent on the clustering of events.

**Data**

*Sumatra–Andaman region*

The USGS PDE data ($m_b \geq 4$) has been used for the period 1996–2005 to study the Sumatra–Andaman region in search of numerical precursors of the 26 December 2004 earthquake. Eighty-two consecutive windows were formed for each of the 50-events period, totalling 4100 earthquakes in the region with latitude (0–10°N) and longitude (92.5–107°E). The analysis was repeated by forming 70 consecutive windows for the 50-events period totalling 3500 earthquakes ($m_b \geq 4$) for the period 1993–2005 in the region with latitude (5°S–5°N) and longitude (92.5–107.5°E). Historical seismicity$^{10}$ formed the basis of selecting our domain of analysis. The 50-events window was taken on the basis of detection of sharp $D_c$ anomaly prior to a large event in a reasonable time-span for proper utility of this warning system in hazard mitigation. If the window size were taken as 100 or 150 events, which is usually done, it will render the detection of the precursor for much longer time-span and less diagnostic. In order to detect a more sensitive temporal variation of $D_c$ within a shorter time-span, it is necessary to analyse time events over a shorter time period as possible$^6$. If we take less than 50-events window, it becomes difficult to maintain the range of correlation coefficients of the regression line for the linear segment, which is considered for more than 0.97 in all the graphs of $\log r$ vs $\log C_d(r)$ used to determine $D_c$ and $D_n$. In general, if this analysis has to be done for an active tectonic setting to find the potential for large earthquakes, one has to start with large spatial window ($10^° \times 10^°$) for 30, 50, 100, etc. events time window and then reanalyse for small spatial window ($5^° \times 5^°$) for 30, 50, 100, etc. events time window until all possible spatio-temporal patterns of seismicity have been found. The range of correlation coefficients of the regression line for the linear segment is considered for more than 0.97 in all the graphs of $\log r$ vs $\log C_d(r)$ used to determine $D_c$. The scaling range for the linear portion of $\log r$ vs $\log C_d(r)$ plot is about 5–50 km, which is well within the region of study considered. The value of scaling region is approximately smaller than 1/3–1/4 of the side length of analysis region, complying with the study$^{11}$ ruling out boundary effects on our analysis.

*Izmit, Turkey region*

The USGS PDE data ($M_d \geq 3$) have been used for the period 1973–2005 for the study of the Izmit region, in search of numerical precursors of the 17 August 1999 earthquake. Sixteen consecutive windows were formed for 50-events period totalling 800 earthquakes in the region with latitude (40–41°N) and longitude (27.6–30.5°E). The analysis was repeated in order to observe precursory condition for different subsets of earthquakes by forming five consecutive windows of 150-events period totalling to 750 earthquakes ($M_d \geq 3$). Also six windows were formed with 30-events period totalling 180 earthquakes ($M_d \geq 4$) in the region with latitude (40–41°N) and longitude (27.6–30.5°E; Figure 2 a and b). The entire analysis indicated that the 50-events set gave a prominent precursor than other subsets. An earlier study$^{12}$ again formed the basis of selecting our domain of analysis.

*Denali Fault, Alaska region*

A similar study was done using USGS PDE data ($m_b \geq 4$) for the period 1973–2005 for the Alaska region, in search of numerical precursors of the 3 November 2002 Denali Fault major earthquake. Thirty-six consecutive windows were formed for 50-events period totalling 1800 earthquakes in the region with latitude (56–66°N) and longitude (138–158°W). An earlier study$^{13}$ formed the basis of selecting our analysis domain.

**Results**

*Sumatra–Andaman region*

The initial study of correlation fractal dimension of all the events with $m_b \geq 4$ occurring in the Sumatra–Andaman
region during 1996–2005 shows that its value varies more or less periodically with time (Figure 3a–e). $D_c$ values have been plotted against mean time of each 50-events window for consecutive periods to study the variation of spatial correlation dimension with time. Clustering was observed for events lying within a region of latitude (0–10°N) and longitude (92.5°–107.5°E) with low $D_c$ values estimated as 1 and 1.17 prior to the 2 November 2002 earthquake of $M_s = 7.6$, and a significantly low $D_c$ value of 0.8 prior to the great Sumatra earthquake of 26 December 2004 with $M_s = 9$ (Figure 1b and c). Teotia and Kumar studied the temporal variation of generalized dimension $D_q$ or $D_q$ spectra for the Sumatra earthquake of 2004. Significant decrease in $D_q$ and $D_o$ spectra was observed prior to the occurrence of the main earthquake using USGS data. The clustering of events was also analysed for regions with latitude (5°S–5°N) and longitude (92.5°–107.5°E) – one with a significantly low $D_c$ value of 1.12 prior to the 2 November 2002 earthquake of $M_s = 7.6$ and the other with a $D_c$ value of 1.18 preceding the 28 March 2005 earthquake of $M_s = 8.7$ (Figure 3d and e). Low $D_c$ value was also observed in the aftershock sequence of the Sumatra earthquake, which can be interpreted by the inherent clustering property of the aftershocks (Figure 3b–e).

Izmit, Turkey region

Similar estimate of spatial correlation dimension and its temporal variations of 50-events window ($M_0 \geq 3$) for consecutive periods from 1973 to 2005 lying between latitude 40°N and 41°N, longitude 26.5°E and 31°E, was analysed for the detection of numerical precursor for the Izmit earthquake of $M_s = 7.8$ (Figure 4a). Significant $D_c$ drops ranging between 0.77 and 0.86 have been observed within several windows with mean time at 28 March 1994, 26 May 1995 and 5 October 1998 prior to the mainshock of 17 August 1999 (Figure 4b).

Denali Fault, Alaska region

Clustering of events ($m_0 \geq 4$) was also noticed during 1973–2005 in the region with latitude (56°–66°N) and longitude (138°–158°W), giving significantly low $D_c$ value of 0.94 occurring prior to the 3 November 2002 Denali earthquake of $M_s = 8.5$ (Figure 5a and b).

Discussion and conclusion

The correlation dimension derived from the above approach reveals that seismic clustering within the subdivisions of the study; the spatial fractal dimension $D_2$ varies from region to region.

Tectonic processes generally activate the fault system where strain accumulation yields a highly stressed zone of asperity. The rupture may nucleate from those weak zones accounting for most of the high frequency seismic energy radiation eventually leading to a large earthquake (Figures 3f, g, 4c and 5c). Recent studies have identified the three distinct asperity–nonasperity zones: an up-dip aseismic zone, a locked seismogenic zone where large slips associated with large rupture would occur followed by an active earthquake release zone at the down-dip section. These zones possess different physical states and properties, and hence can remain elusive from being mapped by various geophysical techniques.

The statistical distribution of low $b$-value, an indicator of the accumulation of stress in the Sumatra region as already reported, finds a significant match with the low $D_c$ distribution by our estimated analysis presented here (Figure 3g).
The combined observational and simulation evidence suggests that the period of increased moment release in moderate earthquakes signals the establishment of long-wavelength correlations in the regional stress field. The central hypothesis in the critical point model for regional seismicity is that it is only during these time periods that a region of the earth's crust is truly in or near a ‘self-organized critical’ (SOC) state, such that small earthquakes participate in cascading into much larger events. This may be due to self-similarity of earthquakes of different scales, which allows the fractures to self-organize in order to attain criticality, as detected by the clustering of events at or in the immediate vicinity of the zone of asperity ultimately causing the mainshock. Clustering can be monitored by the statistical precursor for the major earthquakes by considering the well-constrained earthquake catalogue of seismically active regions of the world. Sammis and co-workers also argued that the observed power-law buildings of intermediate events before a great earthquake represent the approach of the appropriate region towards a state of SOC.

Mignan et al. reported accelerated moment release and stress accumulation model to the Sumatra–Andaman region and also demonstrated that the slip regions of both the 26 December 2004 and 28 March 2005 events have clear evidence of approaching failure prior to both the events, strong evidence of SOC state. More detailed models based on longer period information suggest a concentration in slip and also barrier to rupture that substantiate our low $D_c$ concentrations as well. Thus, in order to study the presence of asperity in the otherwise high seismic regime, the favourable condition for the release of accumulated strain accelerating seismic activity of moderate-sized earthquakes can, therefore, be assessed through the precursory spatio-temporal $D_c$ variation study. This kind of study can be helpful in understanding the cause of great earthquakes as well as the tool for future earthquake forecasting problems by examining the characteristics of $D_c$. Discussions on the advances in statistical analysis of seismological data, and new understanding of the scaling properties of seismicity, including possible universality of major properties of earthquake

Figure 3. (Contd.)

**Figure 3.** a, Temporal distribution of events ($m_i \geq 4$) from 20 July 1996 to 18 September 2005 for latitude (0–10°N), longitude (92.5–107.5°E) plotted against magnitude, where the major events are indicated. b, Temporal variation of $D_c$ where points A (27 July 2000), B (7 January 2002) and C (10 March 2004) represent significant clustering of events before the mainshock of 26 December 2004 ($M_w = 9$) and earthquakes of 2 November 2002 ($M_s = 7.6$) for latitude (0–10°N), longitude (92.5–107.5°E). c, Temporal variation of $D_c$ with the span X and Y representing aftershocks of 26 December 2004 and 28 March 2005 releasing strain energy with significant clustering. d, Temporal variation of $D_c$, where point J (10 December 2001) represents significant clustering of events before the mainshock of 2 November 2002 ($M_s = 7.6$) for latitude (5°S–5°N), longitude (92.5–107.5°E). e, Temporal variation of $D_c$ with the span X and Y representing aftershocks of 26 December 2004 and 28 March 2005 releasing the strain energy with significant clustering, where the point K (10 February 2005) represents significant clustering of events before the mainshock of 28 March 2005 ($M_s = 8.7$) for latitude (5°S–5°N), longitude (92.5–107.5°E). f and g, Spatial distribution of four time windows of 50 events each with low $D_c$ values estimated during the period 8 August 1998–6 September 2004 and one 50-events window during the period 18 February 2005–25 February 2005. The lowest $D_c$ value patch represents the possible asperity or highly stressed region as seen next to the major and Great Sumatra earthquakes.
occurrence provide a unique opportunity to evaluate seismic hazard and estimate the short- and long-term rate of future earthquake occurrence, i.e. to predict earthquakes statistically.
Figure 5.  

(a) Temporal distribution of events ($m_s \geq 4$) from 1973 to 2005 for latitude (56–66°N), longitude (138–158°W). The Denali Fault earthquakes of 23 October 2002 ($M_s = 6.7$) and 3 November 2002 ($M_s = 8.5$) are marked by arrows. 

(b) Temporal variation of $D_C$ presented with significant clustering of events before the mainshock of 23 October 2002 ($M_s = 6.7$) and 3 November 2002 ($M_s = 8.5$) for latitude (56–66°N), longitude (138–158°W) represented by $D_C$ drops. 

(c) Spatial distribution of three time windows of 50 events each with low $D_C$ values obtained during the period 16 November 1997–14 July 1999. The patch of lowest $D_C$ value representing the possible asperity or highly stressed region is seen next to the 3 November 2002 Denali earthquake ($M_s = 8.5$).
The present work corroborates the findings through a b-value study using sliding 50-event time and space window in the Sumatra–Andaman region, wherein two significant b-value drops have been reported, one during the second half of 2002 and the other at the end of 2004 preceding the devastating December 2004 earthquake. What is significant in our findings is that Dc drops to significantly low values at three time windows with specific mean times on 27 July 2000, 1 January 2002 and 10 March 2004, when an analysis was performed on an earthquake catalogue of 4100 events for the period 1996–2005. In the spatial domain, the low Dc concentrations fit well into the low ‘b’ grids indicated. Similarly, correlation between Dc drops and earlier reported b-value decrease was also observed for the Izmith earthquake.


Received 6 February 2007; revised accepted 5 October 2007