united nations educational, scientific and cultural organization

international atomic energy agency

the
abdus salam
international centre for theoretical physics

H4.SMR/1150 - 1

Fifth Workshop on Non-Linear Dynamics and Earthquake Prediction

4 - 22 October 1999

Functions on Catalogs: Measures of Activity, Clustering etc.

I. Rotwain, O. Novikova

International Institute of Earthquake Prediction
Theory and Mathematical Geophysics
Russian Academy of Sciences
Moscow 113556, Russian Federation

Earthquake prediction research is to large extent a trial- and-error procedure. No theory tells us in advance - what are earthquake precursors. So we have to look at the data and find precursors in it. However the data are complex and chaotic. Here, for example is earthquake sequence (Figure 1). This is a real earthquake sequence. How to find precursor here? We have to describe this sequence by robust averaged traits that are useful for prediction. These traits are depicted by functions of time t defined in the sliding time windows with a common end t. One or several of such functions could be used for prediction. And I will describe several functions, which may indicate that a strong earthquake is coming.

I speak about functions which often but not always indicate the approach of strong earthquakes. It is a long way to turn such functions into prediction algorithm. What functions I will show? You know from other lectures that before a strong earthquake the following characteristics of seismicity increase in medium magnitude range (Figure 2):

LEVEL OF SEISMIC ACTIVITY

VARIATION OF SEISMIC ACTIVITY IN TIME

CLUSTERING IN SPACE AND TIME

LONG-RANGE INTERACTION

SPATIAL CONCENTRATION

I will tell you about the functions, which translate these qualitative traits into the quantitative values. These functions can be used in many ways: to compare the seismic situation in the different regions; to compare the real seismicity and the modeled seismicity; and finally to predict earthquakes.

Each function is defined in sliding time window (t-s, t) in some area and magnitude range. We choose them as we wish. For earthquake prediction you may start with areas and magnitudes used in existing algorithms.

FUNCTIONS

The traits we speak about are represented by several not independent functions. One trait can be represented by different functions, or the function with different parameters (different time-window or different range of magnitude). Aftershocks are eliminated, so that relatively strong earthquakes would not dominate earthquake sequence. The number of aftershocks is used as one of the parameters of main shocks. Most of these functions have large values before the strong earthquake.

The following functions (Figure 3) represent the **level of seismic activity:**

$N(t \mid m,s)$ - the number of earthquakes with $M \geq m$ in time interval from (t-s) to t.

$\Sigma(t \mid m,M^I,s,a,b)$ - the weighted number of earthquakes in time interval from (t-s) to t and magnitude ($m \leq Mi \leq M'$).

$$\Sigma\,(t \mid m,M',s,a,b) = \Sigma\, 10^{\,b(M_i-a)}$$

This function can have different sense. If b is about B/3 where B is the coefficient in the relation

$$\lg E = A + BM$$

between the energy E and the magnitude M of an earthquake, then Σ is proportional to the linear size of the earthquake sources. If b is about 2B/3, then the function is proportional to the total area of the sources, and if b is about B, it is proportional to the energy released.

The next function $\mathbf{G(t \mid m_1,m_2,s)}$ - ratio of the number of earthquakes for two magnitude ranges: $m_1 \leq M < m_2$ and $M \geq m_1$. (Figure 4.)

$$G(t \mid m_1, m_2, s) = 1 - \frac{N(t \mid m_2, s)}{N(t \mid m_1, s)}$$

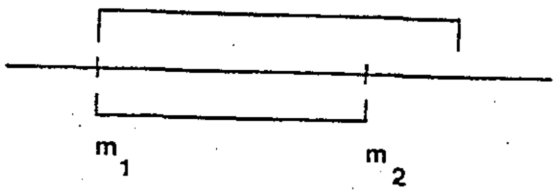

Figure 4.

This function calculated for the time interval (t - s, t).

The next function represents the deficiency of activity i.e. **quiescence** (Figure 5):

$$q(t \mid m,s) = \sum_{+} [a(m) \cdot s - N(t_i \mid m,s)]$$

Here a(m) is an average annual number of earthquakes. $\sum_{+}$ denotes the sum of the positive terms only, that is the sliding time interval (t- s, t), for which the number of earthquakes is less than the average, are only considered. The horizontal **dot** line in the figure indicates the average number N=sa(m) of earthquakes with M ≥ m for s years.

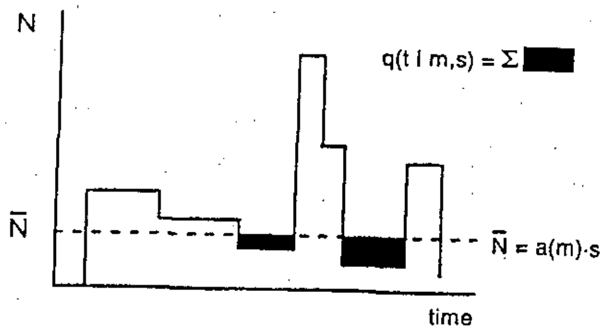

Figure 5.

The following functions represent **variation of seismic activity.**

V(t | m,s,u)-variation of the number of earthquakes (Figure 6). That is a sum of differences between the numbers of earthquakes for two consecutive time intervals (t_{i+1} - s, t_{i+1}) and (t_i - s, t_i), where the moments t belong to the time interval (t– u, t):

$$V(t \mid m,s,u) = \mathrm{Var}\, N(t \mid m,s) = \sum \mid N(t_{i+1} \mid m,s) - N(t_i \mid m,s) \mid$$

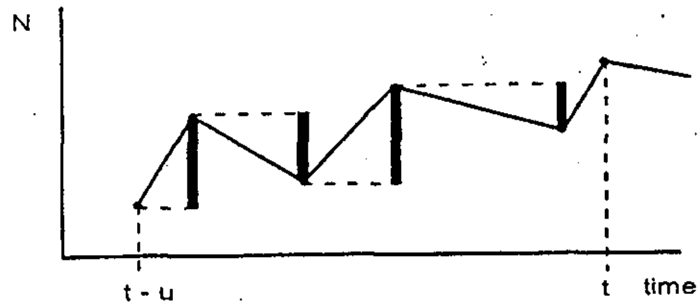

Figure 6.

The next function is **"drop-and-increase" in activity.**
It is the same function as the previous one, but variation of N(t) is counted between t and the previous time of the local maximum of N(t) (Figure 7).

$$Q(t \mid m,s) = \mathrm{Var}\, N(t \mid m,s) = \sum_{I} \mid N(t_{i+1} \mid m,s) - N(t_i \mid m,s) \mid$$

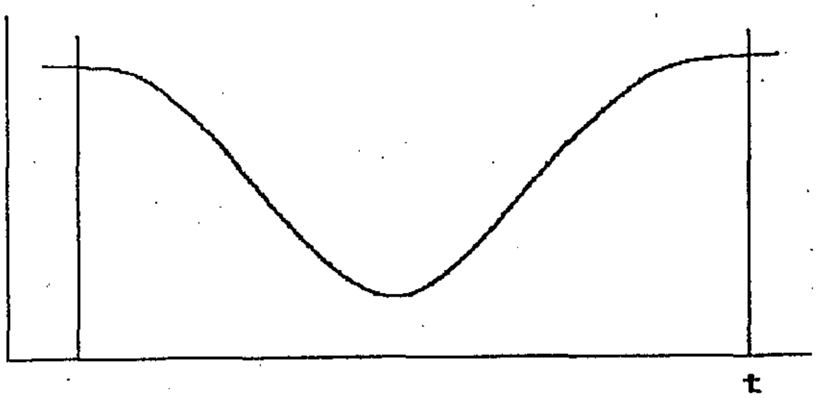

Figure 7.

Here the moments t belong to the time interval (**tmax** , t) where **tmax** is the last moment of the local maximum of N(t) before t.

The next function (Figure 8) is deviation from long-term trend of activity in the period from $\mathbf{t_0}$ to t. Usually $\mathbf{t_0}$ is the beginning of a catalog:

$$L(t|m,s) = N(t|m,t-t_0) - N(t-s|m,t-t_0-s)\frac{t-t_0}{t-s-t_0}$$

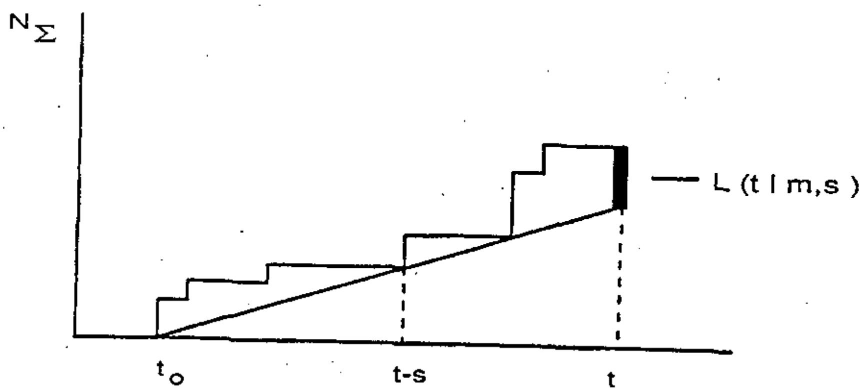

Figure 8.

The second term is a linear extrapolation of N(t) from (t - s) to t.

The next function (Figure 9) represents the increment of activity, that is the difference between the number of earthquakes for two successive time intervals (t - s, t) and (t-2s, t-s):

$$\mathbf{K(t \mid m,s) = N(t \mid m,s) - N(t{-}s \mid m,s)}$$

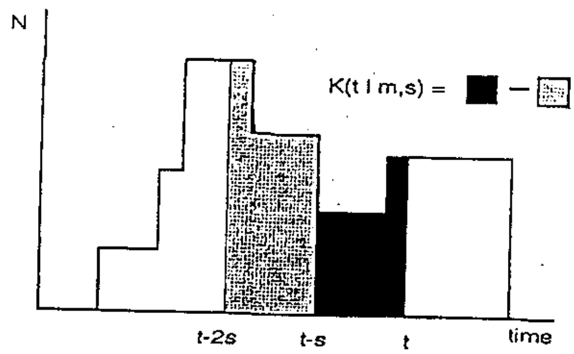

Figure 9.

The next trait of the earthquake flow is **clustering of earthquakes in space and time.** Only one function represents this trait. For each main shock the number **B(e,M)** of aftershocks with magnitude $M \geq M$ in the first e days after the main shock is calculated. The measure of clustering is the maximal B(e,M) for main shocks with magnitude $m \leq M \leq M'$ from the time interval (t - s, t) (Figure 10).

$$\mathbf{B\ (t \mid m,M',s,M\ ,e) = max\ B(e,M\ ).}$$

Many other measures of clustering have been described in the literature. We have chosen the number of aftershocks because to our knowledge it is so far the only precursor for which statistical significance has been established.

The long-range interaction of earthquakes can be characterized by one phenomenon. It was suggested by Prozorov (1982). A strong earthquake is usually followed by some activation in the area where the next strong earthquake is expected. The earthquakes that represent the activation are called long-range aftershocks. Their maximal magnitude is one of the measures of long-range interaction $M_l(t \mid s, M, u)$ (Figure 11).

Here $M_l\ (t \mid s, M_o, u\ )$ - **maximal magnitude of long-range aftershocks in** preceding (t–s,t) **period.**

The **spatial concentration** is represented by two functions.

The first one is roughly proportional to the average area of the source. (Figure 12). Total area of raptures is proportional to

$$\mathbf{\Sigma\ (t \mid m, M', s, a, b) = \Sigma\ 10^{\ b(Mi-a)},\ \ if}$$

$$\mathbf{IgE == A + BM.}$$

$$S = \frac{\sum(t|m, M^{'}, s, \alpha, \beta)}{(N(t|m,s) - N(t|M^{'},s))}$$

where **b = 2B/3. S** is average area of raptures, the maximum of this area within 3 years is considered.

The second one- **Z** -is the ratio of average linear dimension of a source to average distance between them. The maximum of these values within 3 years is considered (Figure 13):

$$Z = \frac{\sum(t|m, M^{'}, s, \alpha, \beta)}{(N(t|m,s) - N(t|M^{'},s))^{2/3}}$$

where **b = B/3.**

The last function represents simultaneous quiescence and activation in adjacent areas. An area is in the quiescence state, if the number of earthquakes for s years is less than some threshold **Nq** and it is in activation state, if the number of earthquakes is greater than the other threshold **Na** (Figure 14). The contrast of activity

Taq(t i m,s,p)

is measured by the time elapsed since the end of the last time period of s years during which the area and its neighbor were in different states.
Na **and** Nq are **(100-p)% and p% percentile** of N respectively.

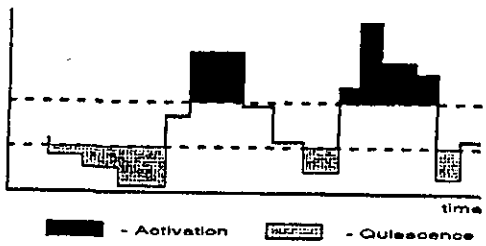

Figure 14.

NORMALIZATION OF FUNCTIONS

Normalization of an earthquake flow is necessary to ensure that a prediction algorithm can be applied with the same set of adjustable parameters in the regions with different seismicity. I tell about normalization by minimal magnitude cut off Mmin, defined by one of two conditions (Figure 15):

Normalization by magnitude thresholds -

a) **M_{min} is determined from condition:** N(M_{min})=b **here** N(M_{min}) **is an average annual number of earthquakes, b is a constant - common for all areas.**

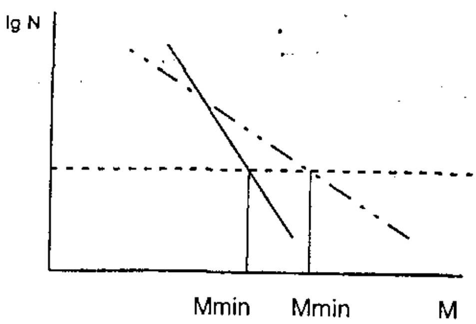

Figure 15.

b) **Mmin = Mo - a; here a is a constant - common for all areas; Mo is the threshold in definition of strong shocks for an area.**

c) Normalization is achieved by the choice of the magnitude range.

- If in a definition of a function the earthquakes are considered with equal weights independent of their magnitudes, then the threshold for M - **Mmin** is defined by the condition: the average annual number of earthquakes with magnitude greater than the threshold equals to the same constant for all regions. This is illustrated in Figure 15.
— If the weights depend on magnitude, the thresholds for M- **Mmin** - depend on the threshold magnitude Mo for strong earthquakes.

How can you use my lecture? First of all these functions may be useful for earthquake prediction in your regions in different combinations. Second, you may experiment to look for similar functions more suitable for your problem. One warning! It should be repeated in many lectures: The functions have many free parameters. That create danger of self-deception. You can always find the function that will predict several earthquakes in your region. The problem is to predict others earthquakes with exactly the same set of parameters. So these functions are useful and powerful instrument but it should be applied with self control. And during the exercises you may compute such functions for your regions.

REFERENCES

1. Keilis-Borok V.I., Rotwain I.M. (1990). Diagnosis of Time of Increased Probability of strong earthquakes in different regions of the world: algorithm CN. Phys.Earth Planet Inter., 61, 57-72.

2. Molchan,G.M., O.E.Dmitrieva, and I.M.Rotwain (1988). Statistical analysis of the results of long term prediction, based on bursts of aftershocks (in Russian), Comput. Seismol, 21, 3-18 (English translation, Allerton, New York).

3. Prozorov.A.G. (1982). Long-range aftershocks, as earthquake precursors in Southern California (in Russian), Comput.Seismol, 14, 20-26 (English translation, Allerton, New York).

4. Zhurkov,S.N., V.S.Kuksenko, V.A.Petrov et al. (1978). Concentration criterion of volume-type fracture in solids (in Russian), in Physical Processes in Earthquake Sources, Nauka, Moscow, 101-116.