Estimates of earthquake recurrences in the Chiayi–Tainan area, Taiwan

Article *in* Engineering Geology · January 2002

DOI: 10.1016/S0013-7952(01)00079-5

Heng Tsai
National Changhua University of Education

ELSEVIER

Engineering Geology 63 (2002) 157–168

ENGINEERING GEOLOGY

www.elsevier.com/locate/enggeo

Estimates of earthquake recurrences in the Chiayi–Tainan area, Taiwan

Heng Tsai*

Department of Geography, National Changhua University of Education, 1, Chinte Road, Changhua, Taiwan

Received 28 March 2000; accepted 8 June 2001

Abstract

In the light of the important role played by the return period in seismic hazard analysis, two types of statistical analyses (Gutenburg–Richter's law and Markov chain) were applied to study the earthquake periodicity in the Chiayi–Tainan area. There are two data sets prepared for this study; one contains earthquakes with $M_L \geq 4$ (283 events) during the 1900–1995 period, the other contains earthquakes with $M_L \geq 2$ (3816 events) during the 1973–1995 period. The mean return period and b values estimated with Gutenburg–Richter relation for earthquakes in the study area are in agreement with previous studies. The mean return period for $M_L \geq 6$ earthquakes is 9.7 years from the data in 1900–1995 period, and 17 years from the data in 1900–1995 period. The Markov chain measures the transition characteristics of earthquakes, and estimates that $M_L = 6–6.9$ and $M_L \geq 7.0$ earthquakes repeat at intervals of 28 and 135 of $M_L \geq 4$ earthquake events, respectively, in the Chiayi–Tainan area. This study also demonstrated that investigation of "substitutability" phenomenon is applicable to earthquake data, which suggests high similarity for low-magnitude earthquakes and dissimilarity for high-magnitude earthquakes. The result from these statistical approaches for earthquake occurrences leads to a conceptual model inferring fault behavior. This model is supported by the recent geologic findings, which may be potentially useful in earthquake hazard assessment.

Keywords: Return period; Gutenburg–Richter's law; Markov chain; Substitutability

1. Introduction

Statistical analyses of seismic records would give an indispensable parameter, i.e., the return period, which enables the seismologist to assess seismic risk. When Gutenburg–Richter's law (denoted as the G–R relation) is applied in calculating seismic periodicity, the recurrence curve offers appealing extrapolations to magnitude levels higher than those included in the data set. Many techniques (Utsu, 1965; Epstein and Lomnitz, 1966; Guttorp, 1987; Lomnitz-Adler and Lomnitz, 1979) have been developed to address the dangers of extrapolating on a log–log plot (the magnitude is itself a logarithmic function). Nevertheless, the straight-line fit made by the least-square method or the maximum likelihood method is still the most commonly used.

A Markov process is a procedure or feature that has an element of randomness or unpredictability, but in which a past event has an influence on a subsequent event. The aspect, in which a subsequent event

* Tel.: +886-4-723-2105x2819; fax: +886-4-721-1186.
E-mail address: geotsaih@cc.ncue.edu.tw (H. Tsai).

0013-7952/02/$ - see front matter © 2002 Published by Elsevier Science B.V.
PII: S0013-7952(01)00079-5

"remembers" a past event, is called a Markov property. Many natural processes (for instance, day-to-day changes in weather) are Markovian. Because the used of the Poisson process to interpret the earthquake occurrences are still debatable (cf. Aki, 1956; Knopoff, 1964; Ferraes, 1967; Isacks et al., 1967; Singh and Sanford, 1972; Udias and Rice, 1975; Smalley et al., 1987; Wang and Kuo, 1998), it seems that the Markov processes might describe the periodic behavior of earthquakes as well.

A Markov chain is the class of Markov processes that exists in discrete times. It has been widely used since Gabriel and Neumann (1962) applied a two-state Markov chain to study the rainfall in Tel Aviv. Harbaugh and Bonham-Carter (1970) discuss the application of Markov chains to simulation studies of geologic processes. For a detailed description about Markov chain theory, see Cox and Miller (1965). The investigation of "substitutability" phenomenon in Markov chain analyses reveals grouping of states that are not apparent otherwise, which has been applied to diverse areas as analysis of parts of speech and automatic processing of satellite photography (Rosenfeld and Huang, 1968). Experiments using data from the Pennsylvanian stratigraphic section in Kansas have been useful in interpreting cyclotherms (Davis and Cocke, 1972).

Since the devastating earthquake in Taiwan on September 21, 1999 (also known as the Chichi earthquake), which caused tremendous life and property losses, people who had witnessed the powerful earthquake began to wonder if a large earthquake would occur in the Chiayi–Tainan area, a highly earthquake-prone sector where population density is high. In this paper, the Markov chain

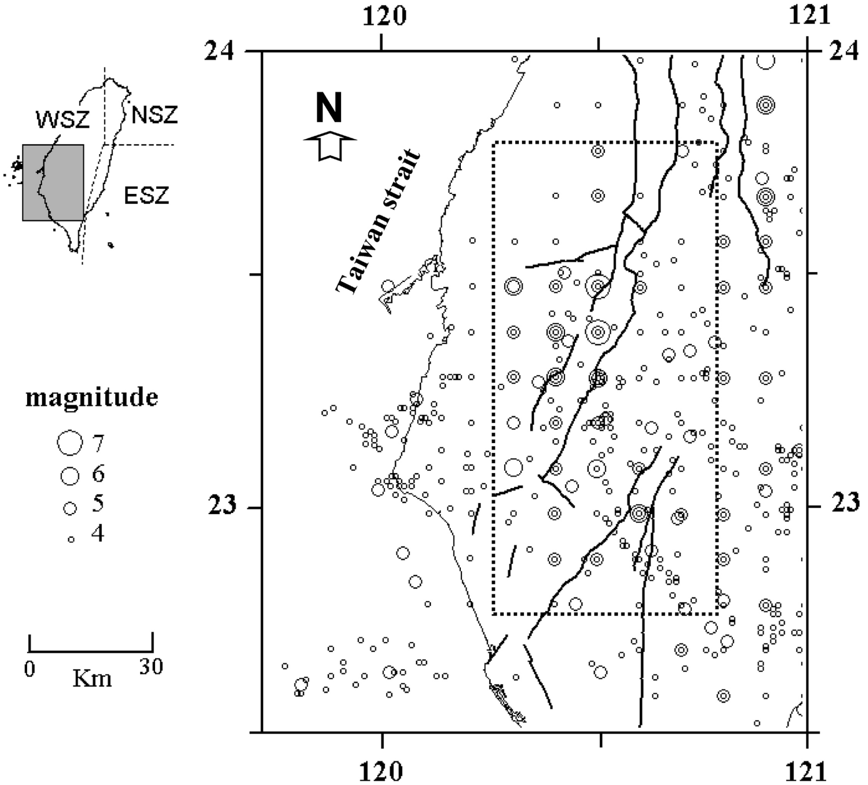

Fig. 1. The $M_L \geq 4.0$ earthquakes occurred in southwestern Taiwan during the 1900–1995 period. Sizes of circles are proportional to earthquake magnitude (see scale). The dotted box indicates the study area, and heavy lines indicate the active faults.

analyses as well as the G–R relation method were performed for seismic periodicity of earthquakes in the Chiayi–Tainan area, Taiwan. Additionally, the substitutability analyses were also performed to study the statistical property of earthquake occurrences.

2. Data

Three seismic belts, known as the western seismic zone (WSZ), eastern seismic zone (ESZ) and north-eastern seismic zone (NSZ), are categorized in accordance with geologic settings, seismicity, and focal mechanisms of earthquakes (Tsai et al., 1977). Numerous damaging earthquakes occurred in the Chiayi–Tainan area of the western seismic zone. The study area is between 22.80°N and 23.80°N (latitude) and 120.25°E and 120.75°E (longitude) (Fig. 1), which is roughly equivalent to zone D defined by Ou (1996) and zone I defined by Cheng et al. (1997). This area is subject to deformation at the front of the Western Foothill Zone. Twelve active faults are known in the area (Chang et al., 1998). Most of the earthquakes that occurred in this area are shallow, no deeper than 70 km in depth.

An earthquake catalog of the area was retrieved from the website of the Institute of Applied Geology, National Central University (Cheng and Lee, 1996). In this catalog, earthquakes are quantified with Richter's local magnitude scale, represented as M_L. In this study, two data sets were prepared; one contains earthquakes with $M_L \geq 4$ (283 events) during the 1900–1995 period, and the other one contains earthquakes with $M_L \geq 2$ (3816 events) during the 1973–1995 period. In the first data set, the earthquakes were classified into four clusters of states, denoted as M_4 (M_L=4–4.9), M_5 (M_L=5–5.9), M_6 (M_L=6–6.9) and M_7 (M_L=7–7.9), respectively. In the second data set, the earthquakes were also classified into four clusters of states, denoted as M_2 (M_L=2–2.9), M_3 (M_L=3–3.9), M_4 (M_L=4–4.9) and M_5 (M_L=5–5.9), respectively. In addition, the author uses the term "forward" to describe the earthquake sequence proceeds through time, and the term "backward" to describe the data sequence conversely from the end of the sequence.

Fig. 2 displays the sequence of $M_L \geq 4$ earthquakes, showing the declining tendency of seismic intensity. Two large earthquakes with M_L=7.1 and M_L=7.2 occurred in 1906 and 1941, respectively. However, the latest two heavily damaging earthquakes (M_L=6.3 and 6.4), which occurred on September 17, 1998 and on October 22, 1999, were not included in the data set. The goodness-of-fit test (Table 1) statistically rejects the hypothesis of the Poisson process for $M_L \geq 5$ earthquakes

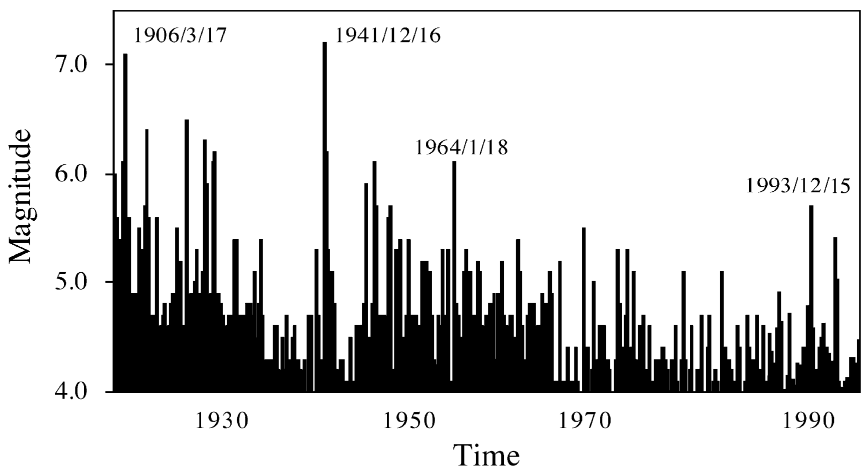

Fig. 2. Time series of the $M_L \geq 4.0$ earthquakes occurring in the study area from 1900 to 1995.

Table 1
Goodness-of-fit test for $M_L \geq 5$ earthquakes

Category	Observed	Prob.	Expected	Contribution to χ^2
0	62	0.50	48	3.904
1	20	0.35	34	5.240
2	9	0.12	11	0.508
≥ 3	5	0.03	3	2.177
	96	1.00	96	11.829

(1) χ^2: Chi-square.
(2) Lambda = 0.688.
(3) degree of freedom (df) = 2.

occurring in the 1900–1995 period at a level of 99% confidence.

3. Gutenberg–Richter relation

The distribution of frequency vs. magnitude of earthquakes in an area is usually described by the Gutenberg–Richter law (Gutenberg and Richter, 1954) as follows:

$$\log N = a - bM \tag{1}$$

where N is the cumulative number of earthquakes with magnitudes greater than M; a and b are constants representing the level of seismic activity and the small-to-large events ratio, respectively. Usually the value of b is in the range of 0.8–1.2 (Evernden, 1970; Wang, 1988), though small regional and temporal differences in b value may be of tectonic significance in themselves.

The b values in Taiwan area have been fully discussed (cf. Tsai et al., 1981; Cheng and Yeh 1989; Wang et al., 1990). Wang (1988) and Cheng and Wang (1985) correlated the b-value distribution with the tectonic setting. In the study area, the plots of $\log N$ vs. M for two time periods are shown in Fig. 3. The straight-line fits were made by least-squares method. The estimated b value is 0.77 for $M_L \geq 4$ earthquakes in the 1900–1995 period and 0.94 for $M_L \geq 2$ earthquakes in the 1973–1995 period.

The G–R relation has a great initial appeal, because it seemingly allows extrapolations to magnitude levels higher than those included in the data set. For instance, one might measure earthquakes of $M_L = 3$ to $M_L = 6$ in a period of only a few years to

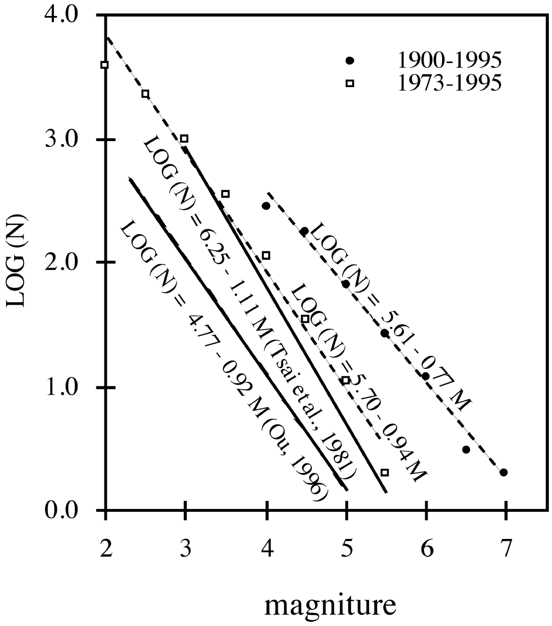

Fig. 3. Plot of $\log N$ vs. magnitude for the study area. Solid circles represent $M_L \geq 4.5$ earthquakes during the 1900–1995 period and open squares represent $M_L \geq 2.5$ earthquakes during the 1973–1995 period. Dotted lines represent the best-fit lines of the data points. Solid and dashed lines represent the G–R relations from Tsai et al. (1981) and Ou (1996), respectively.

delineate the relationship, and then extrapolate the curve as a straight line to $M_L = 8$ to determine how often such an event would occur. The number of the events of specified magnitude level divided by the length of time leads to the mean return period.

The least square method is applied in this study. The mean return periods of earthquakes estimated in this study are in agreement with previous studies (Table 2). The mean return periods of earthquakes

Table 2
Mean return periods estimated based on the G–R relation

	T ($M_L \geq 4$)	T ($M_L \geq 5$)	T ($M_L \geq 6$)	T ($M_L \geq 7$)
Tsai et al. (1981)	0.1	1.4	18	232
Ou (1996)	0.1	0.7	5.6	47
This study[a]	0.2	1.7	9.7	57
This study[b]	0.3	1.9	17	148

T: Return period (years).
[a] Based on the data of 1900–1995.
[b] Based on the data of 1973–1995.

estimated in this study are in agreement with previous studies. For instance, the mean return period for $M_L \geq 6$ earthquakes is 9.7 years from the data in 1990–1995 period, and 17 years from the data in 1973–1995 period. Ou (1996) suggests a period of 5.6 years to reoccur a $M_L \geq 6$ earthquake. Tsai et al. (1981) suggests a period of 18 years for recurrence of earthquake with the same magnitude level. The mean return periods of earthquakes for every magnitude level obtained in this study are between Ou (1996) and Tsai et al. (1981). Calculation from the 1900–1995 data gives smaller return periods than that obtained from the 1973–1995 data. As the magnitude levels increase, the discrepancies of the value of the return period increase.

4. Markov chain analysis

4.1. Theoretical background

Assuming that $X_0, X_1, X_2, X_3, \ldots$ is a time series of random variables with n states, X_i indicates the existing state at time i. Suppose state j exists at time t, then the probability of state k at time $t+s$ shall be:

$$p_{jk}^{(s)} = prob(X_{t+s} = k \mid X_t = j), (t, s = 1, 2, 3, \ldots) \tag{2}$$

where $p_{jk}^{(s)}$ is s-step transition probability, and it means the probability of state $X_{t+s} = k$ (state k exists at time $t+s$) is only associated with the state of $X_t = j$ (the state j exists at time t) and is irrelevant in a state existing at any other time. Such random variables with nature of "memory" in terms of discrete states define the Markov chain. The s-step transitions of states in the system is expressed by the transition probability matrix, P_s, described as follows:

$$P_s = \{p_{jk}^{(s)}\}_{n \times n} \tag{3}$$

Transition matrix has two significant properties: (1) none of the elements $p_{jk}^{(s)}$ is negative ($p_{jk}^{(s)} \geq 0$; $j,k = 1, 2, \ldots, n$); (2) the sum of each row in the matrix equals 1 ($\sum_j p_{jk}^{(s)} = 1$; $j = 1, 2, \ldots, n$). According to the Chapman–Kolmogorv relation that $P_s^{m+n} = P_s^m \cdot P_s^n$, which indicates transition probability at time $m+n$ equals the product of transition probability at time m multiply transition probability at time n. It suggests transition matrix with character of

$$P^m = P_s^{m-1} \cdot P_s \tag{4}$$

where P_s^m represents m-times self-multiplication of transition probability matrix P_s. A fixed unchangeable value obtained among all elements of the matrix indicates a saturated or balanced chain, where the matrix is considered ergodic-state matrix. For a recurrent state, $p_{kk}^{(s)}$ denotes that probability starts at state k and enter state k at time s. According to the limit theorem for Markov chains (Cox and Miller, 1965, p. 96), if k is aperiodic, where a certain state repeats on a non-constant cycle, then

$$\lim_{m \to \infty} p_{kk}^{(s)} = \frac{1}{u_k} \tag{5}$$

where u_k is its mean return period.

4.2. The application

If the occurrence of earthquake is closely related to that of the last one, a Markov chain specified with a one-step transition may be applied to study the seismic periodicity.

Let us take the data of 283 $M_L \geq 4$ earthquakes over 1900–1995 for example: A four-by-four frequency matrix is capable of describing the transition behavior of the sequence. The frequencies of the transition from one state to every other state are tabulated as transition frequency matrix (Table 3). The frequencies in the frequency matrix can be expressed as probabilities by dividing each element of the frequency matrix by the total for its row. For instance, an earthquake of state M_4 was followed by another state M_4 earthquake 177 times (i.e., the number of two earthquakes of $M_L = 4$–4.9 occur successively is 177). But the earthquakes of state M_4 recorded a total of 217 times (row total). Thus, based on the observations, the probability of a state M_4 earthquake occurring immediately after another state M_4 earthquake is $177/217 = 0.816$. The percentage for each row elements of the frequency matrix defines p_{jk} as the probability of a transition from state j to state k, thus forming a forward-transition probability matrix (as shown in Table 3).

Table 3

a. The transition frequency matrix and transition probability matrices for earthquake data in the 1990–1995 period

Transition frequency matrix

from \ to	M_4	M_5	M_6	M_7
M_4	117	37	3	0
M_5	35	13	5	1
M_6	4	3	1	1
M_7	1	0	1	0

Forward-transition probability matrix

from \ to	M_4	M_5	M_6	M_7
M_4	0.186	0.170	0.014	0.000
M_5	0.648	0.240	0.093	0.019
M_6	0.444	0.333	0.111	0.111
M_7	0.500	0.000	0.500	0.000

Backward-transition probability matrix

from \ to	M_4	M_5	M_6	M_7
M_4	0.186	0.698	0.300	0.000
M_5	0.161	0.245	0.500	0.500
M_6	0.018	0.057	0.100	0.500
M_7	0.005	0.000	0.100	0.000

b. The transition frequency matrix and transition probability matrices for earthquake data in the 1973–1995 period

Transition frequency matrix

from \ to	M_2	M_3	M_4	M_5
M_2	2217	548	52	8
M_3	539	293	42	3
M_4	59	36	7	1
M_5	10	0	1	0

Forward-transition probability matrix

from \ to	M_2	M_3	M_4	M_5
M_2	0.785	0.194	0.018	0.003
M_3	0.615	0.334	0.048	0.003
M_4	0.578	0.353	0.069	0.000
M_5	0.909	0.000	0.091	0.000

Backward-transition probability matrix

from \ to	M_2	M_3	M_4	M_5
M_2	0.785	0.625	0.510	0.667
M_3	0.191	0.334	0.412	0.250
M_4	0.021	0.041	0.069	0.083
M_5	0.004	0.000	0.010	0.000

Table 4
The ergodic-state matrices for earthquake data

Period (1990–1995)

$$\begin{array}{cc} & \text{to} \\ & \begin{array}{cccc} M_4 & M_5 & M_6 & M_7 \end{array} \\ \text{from} \begin{array}{c} M_4 \\ M_5 \\ M_6 \\ M_7 \end{array} & \begin{bmatrix} 0.769 & 0.188 & 0.036 & 0.007 \\ 0.769 & 0.188 & 0.036 & 0.007 \\ 0.769 & 0.188 & 0.036 & 0.007 \\ 0.769 & 0.188 & 0.036 & 0.007 \end{bmatrix} \end{array}$$

Period (1973–1995)

$$\begin{array}{cc} & \text{to} \\ & \begin{array}{cccc} M_2 & M_3 & M_4 & M_5 \end{array} \\ \text{from} \begin{array}{c} M_2 \\ M_3 \\ M_4 \\ M_5 \end{array} & \begin{bmatrix} 0.740 & 0.230 & 0.027 & 0.003 \\ 0.740 & 0.230 & 0.027 & 0.003 \\ 0.740 & 0.230 & 0.027 & 0.003 \\ 0.740 & 0.230 & 0.027 & 0.003 \end{bmatrix} \end{array}$$

After 15-times self-multiplication of the forward-transition probability matrix, all elements retain constant values up to sixth decimal places and lead to an ergodic-state matrix. According to Eq. (5), reciprocals of the elements along the main oblique axis in the ergodic-state matrix (Table 4) indicate the mean return periods for each state as follows:

$$M_4 = 1/0.769 = 1.3 \text{ unit},$$

$$M_5 = 1/0.188 = 5.3 \text{ unit},$$

$$M_6 = 1/0.036 = 27.8 \text{ unit},$$

$$M_7 = 1/0.007 = 143 \text{ unit},$$

which the unit is not defined by the length of time but by the interval of events of $M_L \geq 4$ earthquake. As the magnitude level increases, the return period is increased correspondingly. The result indicates that earthquakes of states M_4, M_5, M_6 and M_7 repeat at intervals of 1.3, 5.3, 27.8 and 143 events of $M_L \geq 4$ earthquake, respectively.

Similarly, we perform the calculation for the data in 1973–1995. The ergodic-state matrix (Table 4) is obtained after 17 times of self-multiplication of the forward-transition probability matrix as shown in Table 3—which gives the mean return periods for each state as follows:

$$M_2 = 1/0.740 = 1.4 \text{ unit},$$

$$M_3 = 1/0.230 = 4.3 \text{ unit},$$

$$M_4 = 1/0.027 = 37 \text{ unit},$$

$$M_5 = 1/0.003 = 333 \text{ unit},$$

The result shows that earthquakes of states M_2, M_3, M_4 and M_5 repeat at intervals of 1.4, 4.3, 37 and 333 events of $M_L \geq 2$ earthquake, respectively.

However, Markov chain analysis can give the return period in a form of length of time as well. For instance, if a $M_L \geq 4$ earthquake occurs every 4 months, according to the calculated mean return period from Markov chain analysis, it will take 9.3 years for a M_6 earthquake to reoccur. The analyses reveal that earthquakes of various magnitudes occurred at various rates. In addition, Markov chain analyses measure the transition characteristics of earthquakes, thus they give the interval of events for earthquake recurrence.

5. Substitutability analysis

"Substitutability" is the tendency if two or more states to occur in a common context. From the data sequence, we can develop two quantities: the one-step left substitutability and one-step right substitutability. These terms are derived from problems in the analyses of written text, where the states consist of words, and the strings of states proceed from left to right. If two words tend to be followed by the same words, they can be substituted one another. This is called left substitutability because the word on the left in a sequence is the one that can be replaced by another.

5.1. Forward substitutability

Because earthquake sequence proceeds with time, the term "left substitutability" might more appropriately be considered "forward substitutability". The matrix of forward substitutability is obtained from the forward transition probability matrix by computing a cross product ratio between rows of the matrix.

Based on the previously defined $n \times n$ transition probability matrix, where p_{jk} is the probability that state k follows state j, the cross-product ratio

Table 5

a. The substitutability matrices for earthquake data in the 1990–1995 period

Forward substitutability matrix

$$\begin{array}{cc} & \begin{array}{cccc} M_4 & M_5 & M_6 & M_7 \end{array} \\ \begin{array}{c} M_4 \\ M_5 \\ M_6 \\ M_7 \end{array} & \begin{bmatrix} 1 & 0.98 & 0.87 & 0.70 \\ 0.98 & 1 & 0.94 & 0.75 \\ 0.87 & 0.94 & 1 & 0.54 \\ 0.70 & 0.75 & 0.54 & 1 \end{bmatrix} \end{array}$$

Backward substitutability matrix

$$\begin{array}{cc} & \begin{array}{cccc} M_4 & M_5 & M_6 & M_7 \end{array} \\ \begin{array}{c} M_4 \\ M_5 \\ M_6 \\ M_7 \end{array} & \begin{bmatrix} 1 & 0.99 & 0.66 & 0.15 \\ 0.99 & 1 & 0.76 & 0.29 \\ 0.66 & 0.76 & 1 & 0.71 \\ 0.15 & 0.29 & 0.71 & 1 \end{bmatrix} \end{array}$$

Mutual substitutability matrix

$$\begin{array}{cc} & \begin{array}{cccc} M_4 & M_5 & M_6 & M_7 \end{array} \\ \begin{array}{c} M_4 \\ M_5 \\ M_6 \\ M_7 \end{array} & \begin{bmatrix} 1 & 0.97 & 0.57 & 0.11 \\ 0.97 & 1 & 0.72 & 0.22 \\ 0.57 & 0.72 & 1 & 0.38 \\ 0.11 & 0.22 & 0.38 & 1 \end{bmatrix} \end{array}$$

b. The substitutability matrices for earthquake data in the 1973–1995 period

Forward substitutability matrix

$$\begin{array}{cc} & \begin{array}{cccc} M_2 & M_3 & M_4 & M_5 \end{array} \\ \begin{array}{c} M_2 \\ M_3 \\ M_4 \\ M_5 \end{array} & \begin{bmatrix} 1 & 0.97 & 0.95 & 0.97 \\ 0.97 & 1 & 0.99 & 0.88 \\ 0.95 & 0.99 & 1 & 0.84 \\ 0.97 & 0.88 & 0.84 & 1 \end{bmatrix} \end{array}$$

Backward substitutability matrix

$$\begin{array}{cc} & \begin{array}{cccc} M_2 & M_3 & M_4 & M_5 \end{array} \\ \begin{array}{c} M_2 \\ M_3 \\ M_4 \\ M_5 \end{array} & \begin{bmatrix} 1 & 0.97 & 0.90 & 0.99 \\ 0.97 & 1 & 0.98 & 0.99 \\ 0.90 & 0.98 & 1 & 0.95 \\ 0.99 & 0.99 & 0.95 & 1 \end{bmatrix} \end{array}$$

Mutual substitutability matrix

$$\begin{array}{cc} & \begin{array}{cccc} M_2 & M_3 & M_4 & M_5 \end{array} \\ \begin{array}{c} M_2 \\ M_3 \\ M_4 \\ M_5 \end{array} & \begin{bmatrix} 1 & 0.94 & 0.86 & 0.96 \\ 0.94 & 1 & 0.98 & 0.87 \\ 0.86 & 0.98 & 1 & 0.80 \\ 0.96 & 0.87 & 0.80 & 1 \end{bmatrix} \end{array}$$

defining the forward substitutability between two states is

$$F_{jk} = \frac{\sum_{i=1}^{m} p_{ji} p_{ki}}{\sqrt{\sum_{i=1}^{m} p_{ji}^2 \sum_{i=1}^{m} p_{ki}^2}} \tag{6}$$

Note that if the rows of j and k are identical (i.e., $j = k$), the numerator and denominator become identical and the ratio is $F_{jk} = 1.0$. Therefore, the substitutability measure is constrained in the range $0 \leq F_{jk} \leq 1$. High substitutability indicates a strong tendency for the two states to be followed by similar states. Low substitutability, in contrast, indicates that the two states are usually followed by different states.

The forward substitutability matrix (Table 5a,b) is derived from forward-transition probability matrix in Eq. (6), in which the elements indicate the similarity between one state and another. In the forward substitutability matrix (Table 5a), the substitutability values for the pairs of two states are in the range of 0.98–0.54. The lowest forward substitutability for the pair M_6–M_7 indicates that the two states are usually followed by different states. However, all pairs of two states in the forward substitutability matrix in Table 5b give the value of ≥ 0.84, suggesting high forward substitutability.

5.2. Backward and mutual substitutabilities

Similarly, the matrix of backward substitutability is obtained by computing the cross-product ratios, B_{jk}, between columns of the backward transition probability matrix, which in turn is obtained by dividing each element of the transition frequency matrix by the column totals. This backward transition probability matrix gives the relative frequency with which a state is preceded by another.

In the backward substitutability matrix (Table 5a), the substitutability values for the pairs of two states range from 0.99 to 0.15. The value for the pair M_6–M_7 (0.71) is higher than M_4–M_7 (0.15) and M_5–M_7 (0.29). In Table 5b, all the pairs of two states in the backward substitutability matrix, similar to the forward substitutability matrix, give the value ≥ 0.90.

Furthermore, a matrix of mutual substitutability is obtained by forming the products of all pairs of forward and backward substitutabilities (i.e., $S_{jk} = F_{jk} B_{jk}$). Table 5a shows the values of mutual substitutabilities in the range of 0.97–0.11, similar to those of backward substitutabilities. Three pairs (M_4–M_7, M_5–M_7, and M_6–M_7) suggest low substitutability with their values less than 0.40. However, all pairs of two states in the mutual substitutability matrix in Table 5b give the value ≥ 0.80, suggesting high mutual substitutability.

For better illustration, we can create a "tree" or hierarchical arrangement that shows the relation of one state to another, based on their mutual substitutability, by connecting the states in the order of greatest substitutability. These trees are constructed with the clustering algorithm using distance coefficient (see Chap. 7, Davis, 1973). A low distance indicates two clusters are similar or "close together", whereas a large distance indicates dissimilarity. Fig. 4 shows the results of the hierarchical arrangement depicting the similarity for earthquake clusters. The separation between clusters of M_4 and M_7 is the largest, and that between the clusters of M_4 and M_5 is the smallest of the four clusters of earthquakes in the 1900–1995 period. The clusters of M_4 and M_5 earthquakes are placed together, indicating highest similarity in earthquake sequence. Separation increases as the magnitude level of clusters increases. For earthquakes in the 1973–1995 period, the separation between clusters of M_3 and M_5 is the largest, and that between clusters of

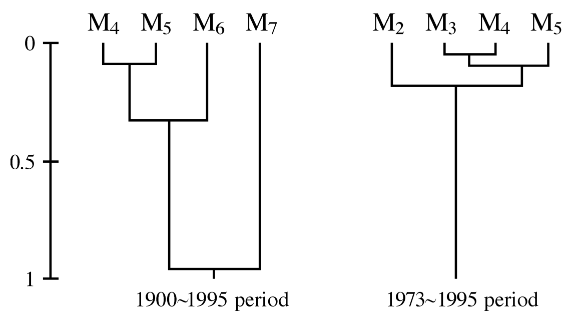

Fig. 4. Hierarchical arrangement for earthquake data. The clusters of earthquakes with the highest mutual similarity are placed together. The left axis shows the values of distance coefficient, which measures from 0 to 1.

M_3 and M_4 is the smallest of the four clusters. However, the low values of distance coefficients for all four clusters indicate their high resemblance. The hierarchical arrangement shows the high similarity for low-magnitude earthquakes and dissimilarity for those high-magnitude earthquakes (say, $M_L = 6$ and $M_L = 7$ earthquakes).

6. Discussion

6.1. The return periods

It is convenient to estimate the return periods of earthquakes by using the G–R relation. However, this study shows a slight difference in data due to area coverage or sample period defined may lead to different return periods for earthquakes. As was demonstrated for southern California by Allen et al. (1965), a detailed 30-year seismographic record for the entire region, including northern Baja California, Mexico, seemed to give meaningful results in estimating how often large events should occur somewhere within the broad region. But when smaller areas and specific faults were examined, the results were considerably less meaningful. Therefore, the return period estimated with the G–R relation can be a valuable source of insight in circumstances when the sample period is long and the area coverage is large. However, the G–R approach falls short of providing information on the transition of earthquakes with different magnitude levels.

Seismic data of the last 10 decades indicates that the seismic activities in the Chiayi–Tainan area are not Poissonian. Wang and Kuo (1998) demonstrated that the series of earthquakes is Poissonian, even though the tectonic settings, the fault distributions and the size of the seismic regions are different. The use of the catalog without removal of foreshocks and aftershocks causes the discrepancies (Gardner and Knopoff, 1974). It, however, implies that a Markov system may be applicable for seismic periodicity. In fact, Markov chain analyses not only estimate the return periods for earthquakes of different magnitude levels but also measures the transition characteristics of earthquakes that occur successively. The probability of each element in the ergodic-state matrices approaches a constant value leading to a mean return period for each state (magnitude level). We can expect similar ergordic Markov chain exist in other seismic zones. The comparison of transition characteristics of earthquakes from place to place will allow us to improve our understanding on the change of stress in an area.

6.2. Model that infers fault behavior

Earthquakes are a natural consequence of the faulting process, especially for the fault system that will repeatedly move, producing stress that exceeds the strength of rocks. When it happens, the rocks fail (rupture) and energy is released. If there is some linkage between time intervals of earthquake events and the amount of stress released in each event, the result of the statistical analyses for earthquake occurrences leads to a conceptual model inferring fault behavior. Fig. 5 shows the pattern of stress in an area that accumulates and releases with time, resulting in a series of earthquake events. The solid line indicates that earthquakes occur after stress to the fault segment has accumulated, and then the stress drop to the base level and magnitude vary. The base level is represented by the dotted line. The stress built up by the long-term barrier from crustal blocks results in an inclined base level. When regional stress reaches its maximum level, the rupture of one fault or fault segment resulting in the ruptures on the adjacent faults or fault segments will trigger an earthquake of high magnitude. In the meantime, the stress fails

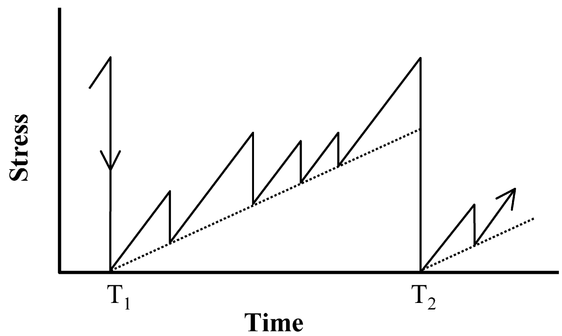

Fig. 5. The conceptual model of earthquake occurrence based on statistical analyses. The figure shows that the pattern of stress drop with time.

back to the ground level. The time interval from T_1 to T_2 (in Fig. 5) defines the return period for this large earthquake to reoccur. A longer time for the accumulation of stress in the crust results in longer return period.

This model suggests that we can estimate the change of regional stress with the recurrence property of high-magnitude earthquakes. The statistical similarity for earthquakes with low magnitude levels reflects the existence and behavior of the fault segments. According to the concept of the model, the return period obtained from Markov chain analyses becomes more meaningful than those obtained from other methods because it measures the transition characteristics of earthquakes. The studies based on paleoseismic trench data indicate that some paleoearthquakes do not rupture the entire segment (McCalpin, 1994; Fumal et al., 1993). Wheeler (1989) shows that persistent segment boundaries control most large ruptures during much of the evolution of the fault. In contrast, nonpersistent boundaries control only a few successive ruptures during a short part of the fault's evolution.

"Earthquake ruptures are termed contagious if the occurrence of the rupture of one fault or fault segment increase the likelihood of the occurrence of rupture on an adjacent fault or fault segment" (Perkins, 1987). It represents the process of one large earthquake influencing the occurrence of another large earthquake on an adjacent or nearby fault. Some earthquakes may result from fault contagion that would be caused by the transmittal of post-earthquake stress changes between adjacent crustal blocks, and would have a lag time of years to centuries (Perkins, 1987).

Furthermore, contagion of faulting should result in a temporal clustering of earthquakes within the fault system, as rupture on one segment induces (after some time) rupture on the adjacent segment. As a result, the temporal clustering is observed, especially in a multi-segment system or in an area where several faults coexist. Fig. 2 shows that most of larger-sized earthquakes occurred almost in the beginning part of the whole time period in consideration, and there is no strong earthquakes with $M_L \geq 6$ occurred after the $M_L = 6.3$ earthquake of 1964 and prior to 1998. With regard to clustering characteristics, we should pay attention to the last two $M_L \geq 6$ earthquakes, which could be the forerunners of a new cluster of larger-sized earthquakes.

7. Conclusions

The b values and the mean return periods estimated with the G–R relation for earthquakes in the Chiayi–Tainan area are in agreement with previous studies. The mean return periods for $M_L \geq 6$ earthquakes is 9.7 years from the data in 1900–1995, and 17 years from the data in 1973–1995. The return period estimated with the G–R relation can be a valuable source of insight during circumstances when the sample period is long and the area coverage is large. However, the G–R approach falls short of providing information on the temporal distribution of earthquakes and the transition of earthquakes with different magnitude levels.

Markov chain analyses show, in addition to the return period giving the length of time, that the transition characteristics of earthquakes are measurable. For instance, $M_L = 6 \sim 6.9$ and $M_L \geq 7$ earthquakes repeat at intervals of 28 and 135 $M_L \geq 4$ earthquake events, respectively, in the Chiayi–Tainan area. Although few published application of substitutability analyses to geologic problems have appeared, this study demonstrated that the investigation of "substitutability" phenomenon is applicable to earthquake data, which suggest high similarity for low-magnitude earthquakes and dissimilarity for high-magnitude earthquakes.

In this study, the statistical analyses for earthquake occurrence leads to a conceptual model inferring fault behavior. This model suggests that the return period obtained from Markov chain analyses is more meaningful than those obtained from other methods. Furthermore, it is supported by recent geologic findings, which may be potentially useful in earthquake hazard assessment.

Acknowledgements

The author is very grateful to the encouragement extended by Dr. P.J. Lee who unfortunately died in November 1999. Thanks are also due to Dr. J.H. Wang and Dr. J.T. Christian for their critical reading and valuable suggestions to improve the author's presentation.

References

Allen, C.R., St. Amand, P., Richter, C.F., Nordquist, J.M., 1965. Relationship between seismicity and geologic structure in the Southern California region. Bull. Seismol. Soc. Am. 55, 753–797.

Aki, K., 1956. Some problem in statistical seismology. Zisin 8, 205–228 [in Japanese].

Chang, H.C., Lin, C.W., Chen, M.M., Lu, S.T., 1998. An introduction to the active faults of Taiwan, explanatory text of the active fault, map of Taiwan scale 1:500000. Cent. Geol. Surv. Spec. Publ. 10. Central Geological Survey, MOEA, Taiwan, R.O.C.

Cheng, C.T., Lee, C.T., 1996. Application of geographic information system to establish seismic database of Taiwan. Annu. Meet. Geol. Soc. China, 583–587 [in Chinese].

Cheng, K.C., Wang, J.H., 1985. The *b*-value distribution and seismicity maps of the Taiwan region. Proc. ROC-JAPAN Joint seminar Multiple Hazards. Natl. Taiwan Univ., Taipei, Taiwan, R.O.C., 69–82.

Cheng, S.N., Yeh, Y.T., 1989. Catalogue of earthquakes in Taiwan from 1604 to 1988. Open-File Rep., Inst. Earth Sci., Acad. Sin., 255 pp.

Cheng, C.T., Lee, C.T., Tsai, Y.B., 1997. Seismic hazard analysis in Taiwan — based on new seismotectonic zones. Ann. Meet. Geol. Soc. China, 300–302 [in Chinese].

Cox, D.R., Miller, H.D., 1965. The Theory of Stochastic Processes. Methuen, London, 308 pp.

Davis, J.C., 1973. Statistics and Data Analysis in Geology. John Wiley & Sons, New York, 550 pp.

Davis, J.C., Cocke, J.M., 1972. Interpretation of complex lithologic successions by substitutability analysis. In: Merriam, D.F. (Ed.), Mathematical Models of Sedimentary Processes. Plenum, New York, pp. 27–52.

Epstein, B., Lomnitz, C., 1966. A model for the occurrence of large earthquakes. Nature 211, 954–956.

Evernden, J.F., 1970. Study of regional seismicity and associated problems. Bull. Seismol. Soc. Am. 60, 393–446.

Ferraes, S.G., 1967. Test of Poisson process for earthquakes in Mexico City. J. Geophys. Res. 72, 3741–3742.

Fumal, T.E., Pezzopane, S.K., Weldon II, R.J., Schwartz, D.P., 1993. A 100-year average recurrence interval for the San Andreas fault at Wrightwood, California. Science 259, 199– 203.

Gabriel, K.R., Neumann, J., 1962. A Markov chain model for daily rainfall occurrence in Tel Aviv. Q. J. R. Meterol. Soc. 88, 90–95.

Gardner, J.K., Knopoff, L., 1974. Is the sequence of earthquakes in Southern California, with aftershocks removal, Poissonian? Bull. Seismol. Soc. Am. 64, 1363–1367.

Gutenburg, B., Richter, C.F., 1954. Seismicity of the Earth and Associated Phenomenon, 2nd edn. Princeton Univ. Press, Princeton.

Guttorp, P., 1987. On least-squares estimation of *b* values. Bull. Seismol. Soc. Am. 77, 2114–2115.

Harbaugh, J.W., Bonham-Carter, G., 1970. Computer Simulation in Geology. Wiley, New York, 575 pp.

Isacks, B.L., Sykes, L.R., Oliver, J., 1967. Spatial and temporal clusters of deep and shallow earthquakes in the Fiji–Tonga–Kermadec region. Bull. Seismol. Soc. Am. 57, 935–958.

Knopoff, L., 1964. Statistics of earthquakes in Southern California. Bull. Seismol. Soc. Am. 54, 1871–1873.

Lomnitz-Adler, J., Lomnitz, C., 1979. A modified form of the Gutenburg–Richter magnitude–frequency relation. Bull. Seismol. Soc. Am. 69, 1209–1214.

McCalpin, J.P., 1994. Quaternary deformation along the East Cache fault zone, Cache County, Utah. Spec. Stud. — Utah Geol. Miner. Surv. 83, 1–37.

Ou, G.B., 1996. A study of seismicity in Tainan–Chiayi area. Proc. 6th Taiwan Symp. Geophys., 57–66 [in Chinese].

Perkins, D.M., 1987. Contagious fault rupture, probabilistic hazard, and contagion observability. In: Crone, A.J., Omdahl, E.M. (Eds.), Directions in Paleoseismology. U. S. Geol. Surv. Open-File Rep., vol. 87-673, pp. 428–439.

Rosenfeld A., Huang, H.K., 1968. An application of cluster detection to test and picture processing. Tech. Rep. 68–68, Computer Science Center, University of Maryland, college Park, MD, 64 pp.

Singh, S., Sanford, A.R., 1972. Statistical analysis of microearthquakes near Socorro, New Mexico. Bull. Seismol. Soc. Am. 62, 917–926.

Smalley, R.F., Chatelain, J.L., Turcotte, D.L., Prevot, R., 1987. A fractal approach to the clustering of earthquakes: applications to the seismicity of the New Hebrides. Bull. Seismol. Soc. Am. 77, 1368–1381.

Tsai, Y.B., Teng, T.L., Chiu, J.M., Liu, H.L., 1977. Tectonic implications of the seismicity in the Taiwan region. Mem. Geol. Soc. China 2, 13–41.

Tsai, Y.B., Liaw, Z.S., Lee, T.Q., 1981. A statistical study of the Taiwan Telemetered Seismographic Network data during 1973–1979. Bull. Inst. Earth Sci., Acad. Sin. 1, 1–22.

Udias, A., Rice, J., 1975. Statistical analysis of microearthquake activity in San Andreas Geophysical observatory, Hollister, California. Bull. Seismol. Soc. Am. 65, 809–827.

Utsu, T., 1965. A method for determining the value of b in a formula log $n = \alpha - bM$ showing the magnitude–frequency relation for earthquakes. Geophys. Bull. 13, 99–103.

Wang, J.H., 1988. *b*-Values of shallow earthquakes in Taiwan. Seismol. Soc. Am. Bull. 78, 1243–1254.

Wang, J.H., Kuo, C.H., 1998. On the frequency distribution of interoccurrence times of earthquakes. J. Seismol. 2, 351–358.

Wang, J.H., Hsieh, C.H., Chan, C.W., Lee, P.H., 1990. Seismicity in the Juisui area of eastern Taiwan. Meteorol. Bull. Cent. Weather Bur. 36, 197–208 [in Chinese].

Wheeler, R.L., 1989. Persistent segment boundaries on basin-range normal faults. In: Schwartz, D.P., Sibson, R.H. (Eds.), Fault Segmentation and Controls of Rupture Initiation and Termination. U. S. Geol. Surv. Open-File Rep., vol. 89-315, pp. 432–444.