

Time-dependent earthquake hazard evaluation in seismogenic systems using mixed Markov Chains: An application to the Japan area

C. Herrera¹, F. A. Nava¹, and C. Lomnitz²

¹CICESE, Seismology Department, Ensenada, B.C., Mexico

²UNAM, Seismology Department, Geophysics Institute, Mexico, D.F.

(Received October 18, 2005; Accepted April 20, 2006; Online published September 16, 2006)

A previous work introduced a new method for seismic hazard evaluation in a *system* (a geographic area with distinct, but related seismogenic regions) based on modeling the transition probabilities of *states* (patterns of presence or absence of seismicity, with magnitude greater or equal to a threshold magnitude M_r , in the regions of the system, during a time interval Δt) as a Markov chain. Application of this *direct* method to the Japan area gave very good results. Given that the most important limitation of the direct method is the relative scarcity of large magnitude events, we decided to explore the possibility that seismicity with magnitude $M \geq M_r^m$ contains information about the future occurrence of earthquakes with $M \geq M_r^m > M_r^M$. This *mixed Markov chain* method estimates the probabilities of occurrence of a system state for $M \geq M_r^M$ on the basis of the observed state for $M \geq M_r^m$ in the previous Δt . Application of the mixed method to the area of Japan gives better hazard estimations than the direct method; in particular for large earthquakes. As part of this study, the problem of performance evaluation of hazard estimation methods is addressed, leading to the use of grading functions.

Key words: Earthquake Hazard, Markov Chains, seismic catalog.

1. Introduction

Seismic hazard assessment is one of the main goals of seismology, because hazard estimates are of primary importance to diminish the social and economical devastating effects of earthquakes. By seismic hazard we mean here the probability of occurrence of earthquakes within a given time, space, and magnitude ranges.

Seismic hazard assessment is often based on statistical analyses of the seismic history of a given area. Many of these assessments are purely probabilistic; some examples are hazard from recurrence-time estimates based on the Gutenberg and Richter distribution (Gutenberg and Richter, 1944), the numerous models based on Poissonian seismicity models (e.g. Brillinger, 1982; Lomnitz and Nava, 1983) the main assumption of these models is that earthquake occurrences are independent in space and time. Nevertheless, modeling the seismicity as a memory-less system may often yield adequate results for areas affected by earthquakes originating in many separate regions, but they are rarely suitable for specific and small areas where the seismic rebound model should apply (Lomnitz and Nava, 1983). For this reason, it is necessary to consider other stochastic models that, in contrast to the Poisson model, describe a type of dependence in a sequence of events, and the Markov model is a method compatible with the factors of causality and memory implicit in the elastic rebound model. Tsapanos and Papadopoulou (1999) used a discrete Markov model in order to model the occurrence of great earthquakes. Vere-

Jones (1966), Knopoff (1971), Hagiwara (1975), and Patwardhan *et al.* (1980) are other examples of the application of the Markov Model. Extreme distributions (Gumbel, 1958) and other empirical distributions (e.g. Epstein and Lomnitz, 1966; Nava and Espindola, 1993) have also been used. Almost all seismic hazard models, including a causal physical component, are based, directly or indirectly, on the elastic-rebound model (Reid, 1910; Richter, 1958). Among these models are those for recurrence times based on seismotectonic arguments, like the time-predictable or slip-predictable models of Shimazaki and Nakata (1980), those based on the seismic gap concept (e.g. Fedotov, 1965; McCann *et al.*, 1979; Kagan and Jackson, 1991), and on seismic migration (e.g. Richter, 1958; Mogi, 1968). Other methods include non-catalog information, such as microseismic, geophysical, or biological precursory data (Aki, 1981; Hagiwara *et al.*, 1997; Yong and Wai, 1995).

There are other statistical techniques based on the detection of seismic spatiotemporal variations (pattern-recognition techniques). For instance, Keilis Borok and Kossobokov (1990) presented a technique for pattern recognition where subjective weights are assigned to the past seismic activity in order to determine times of *increased probability* (TIP) for strong earthquakes. Evidence that information concerning the location and time of occurrence of future large earthquakes can be found in occurrence patterns of smaller earthquakes is on the rise. For example, Agnew and Jones (1991) indicate that at short timescales, some large earthquakes are preceded by smaller earthquakes that occur in zones very close to the source of the great event.

Herrera (2001) and Nava *et al.* (2005) proposed a statistical method for seismic hazard evaluation, based on model-

Copyright © The Society of Geomagnetism and Earth, Planetary and Space Sciences (SGEPSS); The Seismological Society of Japan; The Volcanological Society of Japan; The Geodetic Society of Japan; The Japanese Society for Planetary Sciences; TERRAPUB.

ing the transition probabilities of seismicity patterns, i.e., patterns of occurrence or non-occurrence of large earthquakes in different regions of a given geographic area during a given time interval, as a Markov Chain. Application of this method (henceforth called here the *direct* method) to the Japan area gave extremely good results and suggested some possibilities for future investigations. Since the method presented here is a development of the direct method, we will give a brief summary of its motivation and bases.

2. System Seismic Hazard

The *system* is defined as a geographic area that includes R seismogenic regions. Given a seismic catalog and a starting time, during each successive time interval Δt , the state of each region, s_r , has one of two values: 0 or 1, corresponding, respectively, to the absence or presence within it of earthquakes with magnitude larger than or equal to some threshold value M_r . The total state of the system s is the sum of the regional states:

$$s = \sum_{r=0}^{R-1} 2^r s_r, \quad (1)$$

and there are $S = 2^R$ system states. In binary format, s is simply the concatenation of the binary regional states; it ranges from 00...00 to 11...11 and shows at a glance which regions have earthquakes and which have not for each state. Thus, the *system seismic hazard* is the probability of the system having a given state during a particular interval.

Given that, for a causal system, the stress state of the system (and, hence, the occurrence of large earthquakes in it) depends heavily on the seismic history of each seismogenic region, the possibility of a system having memory and behaving in a Markovian manner was explored.

3. Direct Method

In this method, the state of the system is evaluated for each successive interval, and from this succession of states a transition matrix Θ is constructed with elements θ_{ij} corresponding to the number of times state j occurred after state i . Also constructed is a vector Ξ , where each component $\xi_i = \sum_j \theta_{ij}$ corresponds to the total number of transitions originating from state i . After processing all data, the transition probability matrix P , with elements

$$p_{ij} = \frac{\theta_{ij}}{\xi_i} \quad (2)$$

is built. From the observed state of the system over an interval Δt , the P matrix gives Markovian hazard estimates for each state of the system for the next Δt .

As shown in Nava *et al.* (2005), application of this method to the Japan area with $M_r = 5.5$ gave very good results, much better than those obtained from null hypothesis (memory-less) uniform and Poissonian models. The direct Markovian results had negligible probabilities of being obtained by purely random guessing.

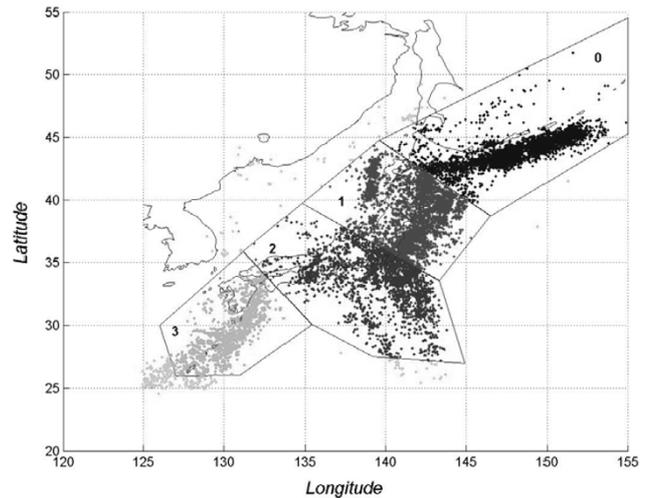


Fig. 1. Seismicity of the study zone reported in the ISC (1964–2002), and the four regions of the system.

4. Mixed Method

Hazard estimates are important mostly for large, potentially dangerous earthquakes, and the most important limitation of the direct method is the relative scarcity of large magnitude events. Given a Δt appropriate for some M_r , when a larger threshold magnitude is considered, then transitions to state 0 (no earthquakes anywhere) dominate, so that other transitions are not very well represented; increasing Δt correspondingly, decreases the number of transitions, which decreases the representativity and robustness of the transition probability matrix.

Hence, we decided to explore the possibility of seismicity with a threshold magnitude M_r^m , having information pertinent to the occurrence of earthquakes with a higher threshold magnitude $M_r^M > M_r^m$. We call this the Mixed Markovian method.

In this method, for each interval Δt , the n th one, say, we have two states of the system s_n^m and s_n^M corresponding to occurrence or non-occurrence of earthquakes with thresholds M_r^m and M_r^M , respectively. Thus, after processing all the data, we have two parallel histories of the system states:

$$s^m = \{s_n^m; n = 1, \dots, N\} \quad \text{and} \\ s^M = \{s_n^M; n = 1, \dots, N\},$$

where N is the total number of time intervals used for constructing the matrix.

From these two system-state histories, we construct a transition matrix Θ , with elements θ_{iJ} corresponding to the number of times state $s_{n+1}^M = J$ occurred after state $s_n^m = i$ (we will use lowercase letters for states with threshold M_r^m , and uppercase ones for states with threshold M_r^M). After processing all data, we construct the transition probability matrix P , with elements

$$p_{iJ} = \frac{\theta_{iJ}}{\xi_i}, \quad (3)$$

where $\xi_i = \sum_J \theta_{iJ}$ corresponds to the total number of transitions originating from state i .

Table 1. System states: decimal (left) and binary (right) showing which regional states (0 or 1) it comprises.

| Status | Region | | | | Status | Region | | | |
|--------|--------|---|---|---|--------|--------|---|---|---|
| | 3 | 2 | 1 | 0 | | 3 | 2 | 1 | 0 |
| 0 | 0 | 0 | 0 | 0 | 8 | 1 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 1 | 9 | 1 | 0 | 0 | 1 |
| 2 | 0 | 0 | 1 | 0 | 10 | 1 | 0 | 1 | 0 |
| 3 | 0 | 0 | 1 | 1 | 11 | 1 | 0 | 1 | 1 |
| 4 | 0 | 1 | 0 | 0 | 12 | 1 | 1 | 0 | 0 |
| 5 | 0 | 1 | 0 | 1 | 13 | 1 | 1 | 0 | 1 |
| 6 | 0 | 1 | 1 | 0 | 14 | 1 | 1 | 1 | 0 |
| 7 | 0 | 1 | 1 | 1 | 15 | 1 | 1 | 1 | 1 |

Before applying this method to the same data base used for the direct method and comparing results, an appropriate method for performance of evaluation needs to be devised.

5. Performance Evaluation

For any method of hazard estimation, one direct measure of its performance is assessment of the probability it assigned to the actual outcome, or outcomes. For our case, when before transition n the system is in state i , the probabilities p_{iJ} ; $J = 1, \dots, R$, constitute the hazard estimates for the next state; if state k occurs, then the *observed transition probability* is $\hat{p}_n = p_{ij}$. The mean observed probability

$$\hat{p} \equiv \frac{\sum_{n=1}^{n_t} \hat{p}_n}{n_t}, \tag{4}$$

where n_t is the number of realized transitions, is a measure of the average probability with which actually occurring states were expected. According to this measure, the best of several hazard estimate methods would be the one yielding the highest \hat{p} , and to judge whether a given hazard estimate by itself is good, its \hat{p} can be measured against the “natural” reference level, which is the maximum entropy probability corresponding to the null hypothesis uniform probability where all states are equally likely to occur:

$$p_{ij}^U \equiv u = S^{-1} = \langle p_{ij} \rangle. \tag{5}$$

However, the \hat{p} measure does not take into account that the main object of hazard estimation is to *forecast* earthquake occurrences so that society may be prepared for them. A useful evaluation must take into account factors like the probability level of forecasts, multiple (and contradictory) forecasts, false alarms, missed forecasts, etc. We will now define quantitative measures for results that characterize the performance of a given method or model.

We define a *forecast* as the statement that a given outcome has a *high* probability of occurring; where a probability is considered high if it lies above a given threshold

$$p_x = f_x u, \tag{6}$$

where f_x is a *success factor* which expresses the threshold in terms of the uniform probability u . A forecast is *successful* when an outcome with $p_{ij} > p_x$ occurs, and the number of successful forecasts is n_x . When a forecast is

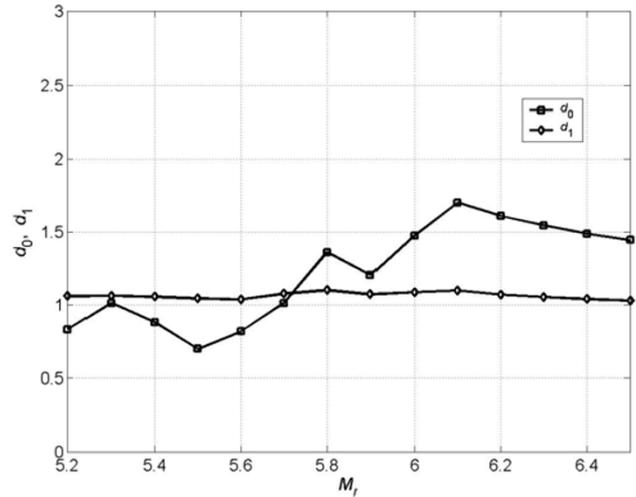


Fig. 2. Grades of the direct method as a function of the threshold magnitudes for *aftcasts*.

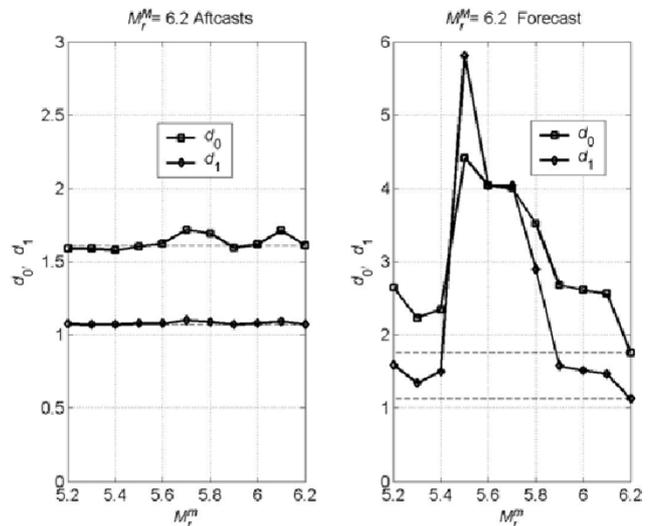


Fig. 3. Optimal grades for the mixed method with $M_f^M = 6.2$ as a function of M_f^m . Horizontal dashed lines are best values for direct method.

not successful, then it is a *false alarm* (type I error), and the number of false alarms will be denoted by n_f .

The *multiplicity*, m_i , is the number of elements larger than p_x in row i of the transition probability matrix, i.e. the number of simultaneous forecasts. When $m_i = 0$, then there is no forecast (no success or false alarm) and outcome i is a *missed event* (type II error). The number of missed events is $n_s = n_t - n_p$, where n_p is the number of transitions for which there was a forecast and n_t (as defined above) is the total number of considered transitions. In order to be very strict with successes, each forecast success was divided by the corresponding multiplicity; for example, for $m_i = 2$, a successful forecast would count as half a success plus one false alarm, while an unsuccessful one would count as two false alarms.

The *regional error*, e , is the number of regions whose activity (occurrence or nonoccurrence of earthquakes) was erroneously forecast.

All the above counts are normalized by n_t so that perfor-

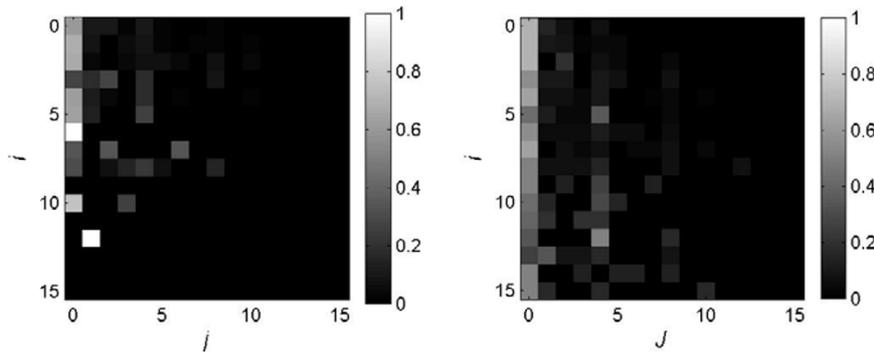


Fig. 4. Transition probability matrices: direct method with $M_r = 6.2$ (left), and mixed method for $M_r^M = 6.2$ and $M_r^m = 5.7$ (right).

mances with different lengths may be compared.

There is one last factor to consider, and it is the usefulness of the forecasts; i.e. to have non-trivial forecasts. There are two trivial cases: one is when the threshold magnitude is larger than the largest observed one; in this case, the probability of no earthquakes at all $p_{00} = 1$, and all other probabilities are null. The second trivial case is when the threshold magnitude is very small and $p_{S-1S-1} = 1$. This case is not so important to us since we are interested in large earthquakes only. For both trivial cases, the measures described above would yield optimum values, but the forecasts would be completely useless because they would carry no information at all. Thus, to avoid trivial, or close to trivial, cases it is necessary to penalize a hyperabundance of transitions ending in the 0 and $S - 1$ states; we will do this by considering the diminishing information content of the corresponding transition probabilities.

The information (in bits) contained in a forecast with probability p is commonly defined as $I(p) = -\log_2(p)$ (cf. Goldman, 1953). Here, we will use ubit information units that assign a value of 1 to the information in the uniform probability, i.e. $I(u) = 1$ ubit. If a total of N_t transitions have been used to evaluate the transition probabilities, then the probability of any transition ending in state 0 is

$$p_0 = \frac{\sum_i p_{i0}}{N_t}, \tag{7}$$

and, for $p_0 > u$, all \hat{p}_{i0} probabilities and forecast successes are multiplied by $I(p_0)$ (in ubits) before being counted. Observed \hat{p}_{iS-1} probabilities and successes are qualified in a similar way.

All measures except \hat{p} depend on the choice of success factor; a large f_x is desirable because we want forecasts to be made for high probabilities, and it will minimize false alarms, but too large a value can so reduce the number of forecasts (increasing the number of missed events) and, hence, of successes, as to make the model almost useless. A low f_x will result in high multiplicity and yield a large number of false alarms and increase the regional error (both undesirable). So, the optimum value of f_x has to be found in order to get the best performance out of a given model (a given combination of system, Δt , M_r^m , and M_r^M ; for the direct method $M_r^m = M_r^M = M_r$).

Choosing the method or model which yields the best performance is not straightforward, because there are usually

tradeoffs between desirable and undesirable traits, which make direct inspection and comparison unpractical. Therefore, we decided to make use of grading functions, i.e. mathematical functions which take into account all relevant factors, weighted according to their relative size and importance, and combines them in such a way that desirable features increase its value and undesirable ones decrease it. Adjusting for relative size is necessary because we may be comparing quantities with different orders of magnitude (e.g. average probability vs. normalized number of successes or false alarms). Weighting for importance is largely subjective, but reflects a consensus of desirability for different traits; for instance, false alarms are quite undesirable (for many obvious reasons) and it is usually preferable to have fewer false alarms than more successes.

There is no rule to say which form a grading function should take. We tried linear, non-linear, product, and mixed grading functions, but in this work we show the results of only two:

$$d_0 = 0.8 + 5\hat{p} + \left(\frac{10n_x - n_f - e - n_s}{n_t} \right) + 0.00001 f_x, \tag{8}$$

$$d_1 = 1.0 + \left(\frac{(400 + 0.00002 f_x) \hat{p}^2 n_x^2}{n_t(n_f + e + n_s)} \right). \tag{9}$$

Of course, the absolute values from a grading function are quite arbitrary and can be changed by modifying some of the baseline or scaling parameters (chosen here so both grades can be clearly seen when plotted using the same scales); but we are mostly interested in the relative values, or relative optimum values (using always the same parameters in the grading function, of course). However, some idea of what the actual values do represent can be had by comparison with grades obtained for the reference “null hypothesis” Poisson and uniform (mere guess) models.

In actual forecasting, all available data to date are used to estimate the transition probabilities used in forecasting the system state for the time interval beginning immediately. A serious problem is that a minimum number of transitions must be used to achieve robust probability estimates, and this may not leave enough transitions to have a representative sample for assessing the model’s performance if the catalog is not long enough. This was the case for our application discussed below; a stopgap solution we adopted, until enough forecast transitions are observed, is to *aftcast* all

Table 2. Best results of state aftcasting. n_t is the number of transitions, f_x is the success factor, n_x is the number of successes, p^b is the Bernoulli binomial probability of observing n_x successes in n_t transitions with uniform probability u ; d_0 and d_1 are grades from the grading functions.

| Method | M_r | n_t | f_x | n_x | d_0 | d_1 | p^b |
|--------|-------------------------------|-------|-------|-------|-------|-------|----------------------|
| Direct | $M_r \geq 6.1$ | 384 | 6.2 | 49 | 1.699 | 1.100 | 1.2×10^{-6} |
| Mixed | $M_r^M = 6.1$ & $M_r^m = 5.7$ | 384 | 5.0 | 51 | 1.661 | 1.095 | 2.3×10^{-7} |
| Mixed | $M_r^M = 6.2$ & $M_r^m = 5.7$ | 384 | 5.0 | 51 | 1.717 | 1.100 | 2.3×10^{-7} |
| Mixed | $M_r^M = 6.3$ & $M_r^m = 5.7$ | 384 | 6.2 | 43 | 1.680 | 1.077 | 8.9×10^{-5} |
| Mixed | $M_r^M = 6.4$ & $M_r^m = 5.8$ | 384 | 7 | 37 | 1.607 | 1.058 | 2.7×10^{-3} |
| Mixed | $M_r^M = 6.5$ & $M_r^m = 6.1$ | 384 | 6.5 | 33 | 1.525 | 1.039 | 1.4×10^{-2} |

Table 3. Best results of state forecasting. Quantities are the same as in Table 2.

| Method | M_r | n_t | f_x | n_x | d_0 | d_1 | p^b |
|--------|-------------------------------|-------|-------|-------|-------|-------|----------------------|
| Direct | $M_r \geq 6.1$ | 20 | 7.5 | 4 | 2.794 | 1.632 | 2.6×10^{-2} |
| Mixed | $M_r^M = 6.1$ & $M_r^m = 5.5$ | 20 | 7.5 | 6 | 4.680 | 6.686 | 9.4×10^{-4} |
| Mixed | $M_r^M = 6.2$ & $M_r^m = 5.5$ | 20 | 5 | 6 | 4.420 | 5.804 | 9.4×10^{-4} |
| Mixed | $M_r^M = 6.3$ & $M_r^m = 5.8$ | 20 | 7.5 | 4 | 3.327 | 2.349 | 2.6×10^{-2} |
| Mixed | $M_r^M = 6.4$ & $M_r^m = 5.8$ | 20 | 7.5 | 4 | 3.327 | 2.349 | 2.6×10^{-2} |
| Mixed | $M_r^M = 6.5$ & $M_r^m = 5.8$ | 20 | 7.5 | 4 | 3.457 | 2.995 | 2.6×10^{-2} |

384 available transitions, i.e. do “forecasts” for data already used in estimating the probabilities, and compare these results with the true forecasts of 20 transitions (transition number 365 was made from probabilities estimated using the first 364 transitions; number 366 using transitions up to 365, and so on). As will be shown, aftcast and forecast performance evaluations roughly agree; but comparisons will be based mainly on aftcast results.

We use the whole catalog, without eliminating aftershocks, for the following reasons. First, most aftershocks occur close in time after the main shock so that they have a strong probability of being part and parcel of the activity within the time interval in which the main event occurred, in which case their presence or absence is immaterial. Second, aftershocks smaller than the threshold magnitude (the great majority) are discarded anyway. Third, if large aftershocks do occur during the following interval in a large number of cases, we want to be able to forecast them. If aftershocks were the main constituent of the system’s memory, then the diagonal elements would dominate the transition probability matrix; which is clearly not the case (Fig. 4).

6. Application and Results

The models described in the sections II and III were applied to the Japan area, using data, from 1964 to 2002, from the International Seismological Centre. The system was chosen because of its tectonically distinct seismogenic regions, high seismicity (Fig. 1), and good catalog coverage. Four regions, shown in Fig. 1, were chosen: Kurile Islands, Central Japan, SE Japan, and Ryukyu Islands, defining 16 system states (Table 1). Nava *et al.* (2005) applied the direct method to a combination of $\Delta t = 0.10$ yr and $M_r = 5.5$. We have kept the same time interval (384 transitions), and explored models with threshold magnitudes $M_r^m = 5.7$ and

$M_r^M = 6.2$. The following figures show results obtained from our analysis.

Based on the values of the grading functions, we found the optimal f_x for each model, and these optimal models were then compared to see both how the mixed method compared with the direct one and which would be the best model to use for issuing useful forecasts.

Figure 2 shows the aftcast performance of the direct method, indicating that best results are obtained for $M_r = 6.1$ ($f_x = 6.2$), with a secondary peak at $M_r = 5.8$ ($f_x = 5$). Forecast grades (not shown) have a peak at $M_r = 5.8$ ($f_x = 4.5$), but have smaller values at higher magnitudes, and grow for smaller ones.

The mixed method for $M_r^M = 6.1$, does not improve the grades, although it does give a larger amount of successes (Table 2). The main motivation behind the mixed method is to estimate seismic hazard for higher magnitudes, and the results shown in Table 2 show that for higher M_r^M , the mixed method does considerably improve the grades (compare values with Fig. 2). The last column of the table is the probability of obtaining n_x successes in n_t transitions by chance given a uniform probability u ; it can be seen that the direct method compares most favorably with chance forecasting, but the mixed method compares even better.

For forecasts, the mixed method for $M_r^M = 6.1$ does give better results, in all aspects, than the direct method (although 20 transitions do not constitute a really significant sample); the comparative results are shown in Table 3.

Figure 3 shows the aftcast grades of the mixed method for $M_r^M = 6.2$, as a function of M_r^m ; the horizontal dashed lines indicate the (optimum) values for the direct method. It can be clearly seen that the mixed method gives better results than the direct method for several values of M_r^m ; both grading functions show improvements for $M_r^m = 6.1$,

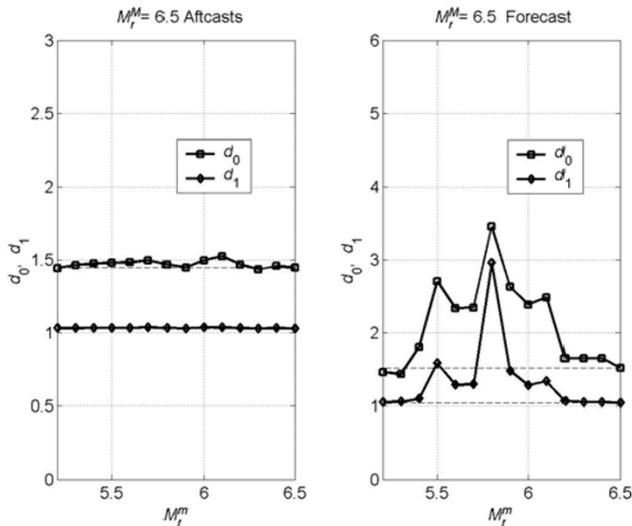


Fig. 5. Optimum grades for the mixed method with $M_r^M = 6.5$ as a function of M_r^m . Horizontal dashed lines are best values for direct method.

6.0, 5.8, 5.7, and 5.6, with the maximum at 5.7. The mixed method not only improves on the direct one for $M_r^M = M_r = 6.2$, but gives even better results than those of the direct method for $M_r = 6.1$ (Table 2). Results are even better for forecasts (Fig. 3 and Table 3).

Figure 4 shows the gray scale-coded transition probability matrices P obtained by the direct method with $M_r = 6.2$ (left), and with the mixed method for $M_r^M = 6.2$ and $M_r^m = 5.7$ (right). It can be seen that, although the overall shape is the same, maxima differ between them; the matrix for the mixed method is not a scaled version of the direct one.

The ratio $r = (d_0^{(\text{mixed})} - d_0^{(\text{direct})})/d_0^{(\text{direct})}$ for *aftcasts* of the whole catalog (384 transitions) and forecasts of 20 transitions, shows that the mixed method gives hazard estimations 6.6% better than the estimated by the direct method (for aftcast and threshold magnitudes $M_r = 6.2$ and $M_r^M = 6.2$). For forecasts and threshold magnitudes $M_r = 6.1$ and $M_r^M = 6.1$, r indicates an improvement of 67% in the hazard estimations of the mixed method. Very similar values result for the corresponding d_1 estimates.

For larger M_r^M , the mixed method consistently has better performance, although with diminishing returns, than the direct method all the way to $M_r^M = 6.5$ (Tables 2 and 3), after which magnitude the Markovian model results lack information content; Fig. 5 shows the behavior of grades of this border case. For all cases, forecasts yield higher determinant values. For actual forecasting at a given target (higher) threshold, the lower threshold used should be the one giving the best results, those with higher grades, for the forecast history up to the current date. How large can the higher threshold be is determined by the length of the homogenous catalog.

7. Conclusions

Application of both Markovian methods to the Japan area, for *aftcasts* of the whole catalog and *forecasts* of 20 transitions, yields extremely satisfactory results that have

negligible probabilities of being obtained by purely random guessing or by a memory-less model; a Poissonian model yields grades about 20% smaller than those obtained by Markovian methods. This result confirms the results of Nava *et al.* (2005) that indicate that this seismic system is not memory-less.

For the direct method, use of grading functions suggests that, for time intervals of $\Delta t = 0.10$ year, best results are obtained when predicting activity with magnitudes $M \geq M_r = 6.1$; these results are better than those for $M_r = 5.5$ reported in Nava *et al.* (2005), and this is good because larger magnitudes are of greater importance for seismic hazard assessments.

The mixed method consistently gives consistently better results than the direct one, particularly for true forecasts (although, unfortunately, there are not enough cases to be statistically significant). This result a very important implication: there is some information about future earthquakes with $M \geq M_r^M$ in the occurrence of slightly smaller ones $M \geq M_r^M > M_r^m$. A practical corollary is that it is possible to extend the forecasting magnitude limit of a given catalog, to estimate seismic hazard for slightly larger, more important, events.

Acknowledgments. Our thanks to Ewa Glowacka, Cecilio Rebollos, Luis Munguía, and Javier Helenes for helpful comments and criticisms. Our appreciation to the anonymous reviewers for their helpful, constructive comments. This research was partially funded by the CONACYT scholarship No. 142091 (C.H.).

References

- Agnew, D. D. and L. M. Jones, Prediction probabilities from foreshocks, *J. Geophys. Res.*, **96**, 11959–11971, 1991.
- Aki, K., A probabilistic synthesis of precursory phenomena, in *Earthquake Prediction, an International Review*, edited by D. W. Simpson and P. G. Richards, 680 pp., M. Ewing series, v. 4, Am. Geophys. Union, Washington, D.C., 1981.
- Brillinger, D., Seismic risk assessment: some statistical aspects, *Earthq. Predict. Res.*, **1** 183–195, 1982.
- Epstein, B. and C. Lomnitz, A model for the occurrence of large earthquakes, *Nature*, **211**, 954–956, 1966.
- Fedotov, S., Regularities of the distribution of strong earthquakes in Kamchatka, the Kurile Islands, and northeast Japan, *Trudy Inst. Fiz. Zemli. Acad. Nauk. SSSR.*, **36**, 66–94, 1965.
- Goldman, S., *Information Theory*, 385 pp., Dover Publ. Inc., USA, 1953.
- Gumbel, E. J., *Statistics of Extremes*, 375 pp., Columbia University Press, New York, N.Y., 1958.
- Gutenberg, B. and C. Richter, Frequency of earthquakes in California, *Bull. Seism. Soc. Am.*, **34**, 185–188, 1944.
- Hagiwara, Y., A stochastic model of earthquake occurrence and the accompanying horizontal land deformations, *Tectonophysics*, **26**, 91–101, 1975.
- Hagiwara, Y., H. Tajima, S. Izutnya, and H. Hanada, Gravity changes associated with earthquake swarm activities in the eastern part of Izu peninsula, *Bull. Earthq. Res. Inst. Tokyo Univ.*, **52**, 141–150, 1997.
- Herrera, C., Determinación de peligro sísmico mediante cadenas de Markov, MSc Thesis, CICESE, 86 p., 2001.
- International Seismological Center, On-line Bulletin, <http://www.isc.ac.uk/Bull>.
- Kagan, Y. and D. Jackson, Seismic gap hypothesis: Ten years after, *J. Geophys. Res.*, **96**, 21419–21431, 1991.
- Keilis-Borok, V. I. and V. G. Kossobokov, Premonitory activation of earthquake flow: algorithm M8, *Phys. Earth. Planet. Int.*, **61**, 73–83, 1990.
- Knopoff, L., A stochastic model for the occurrence of main sequence earthquakes, *J. Geophys. Res.*, **75**, 5745–5756, 1971.
- Lomnitz, C. and F. Nava, The predictive power of seismic gaps, *Bull. Seism. Soc. Am.*, **73**, 1815–1824, 1983.
- McCann, W., S. Nishenko, L. Sykes, and J. Krause, Seismic gaps and plate

- tectonics: seismic potential for major boundaries, *Pure Appl. Geophys.*, **117**, 1082–1147, 1979.
- Mogi, K., Migration of Seismic Activity, *Bull. Earthq. Res. Inst.*, **46**, 53–74, 1968.
- Nava, F. A. and V. Espíndola, Seismic hazard assessment through semi-stochastic simulation, *Bull. Seism. Soc. Am.*, **83**, 450–468, 1993.
- Nava, F. A., C. Herrera, J. Frez, and E. Glowacka, Seismic hazard evaluation using Markov chains; Application to the Japan area, *Pure Appl. Geophys.*, **162**, 1347–1366, 2005.
- Patwardhan, A. S., R. B. Kulkarni, and D. Tocher, A semi Markov model for characterizing recurrence of great earthquakes, *Bull. Seism. Soc. Am.*, **70**, 323–347, 1980.
- Reid, H. F., The mechanism of the earthquake, The California earthquake of April 18, 1906, Report of the Research Senatorial Commission, Carnegie Institution, Washington D.C., 2, 16–18, 1910.
- Richter, C., *Elementary Seismology*, 521 pp., W. H. Freeman, San Francisco, 1958.
- Shimazaki, K. and T. Nakata, Time predictable recurrence model for large earthquakes, *Geophys. Res. Lett.*, **7**, 279–282, 1980.
- Tsapanos, T. and A. A. Papadopoulou, A discrete Markov Model for earthquake occurrences in Southern Alaska and Aleutian Islands, *J. Balkan Geophys. Soc.*, **2**(3), 75–83, 1999.
- Vere-Jones, D., A Markov model for aftershock occurrence, *Pure Appl. Geophys.*, **64**, 31–42, 1966.
- Yong, S. and Z. Wai, The correlation between random variation and Earth solid tide Change in Rock-Groundwater system-the Mechanical Foundation for Using Change to predict Earthquake, *J. Earthq. Predict. Res.*, **4**, 3–15, 1995.
-
- C. Herrera (e-mail: cherrera@cicese.mx), F. A. Nava (e-mail: fnava@cicese.mx), and C. Lomnitz (e-mail: cinna@prodigy.net.mx)