The precursory earthquake swarm in Japan: hypothesis test

F. F. Evison[1] and D. A. Rhoades[2]

[1]*Institute of Geophysics, Victoria University of Wellington, P.O. Box 600, Wellington, New Zealand*
[2]*Institute of Geological and Nuclear Sciences, P.O. Box 30-368, Lower Hutt, New Zealand*

A formal performance test of the hypothesis that swarms are long-term precursors to major shallow earthquakes is in progress in the region of Pacific-plate subduction in Japan. The likelihood of the major earthquakes (JMA magnitude $\geq$6.8) that occurred in the region during the test period (1983–1998) was 5.04 times higher under the swarm hypothesis than under the stationary Poisson model; this result is inconclusive in terms of the proposed acceptance level of 20. The earthquakes were the Hokkaido-Toho-Oki earthquake (M8.1) of 4 October 1994, and the Sanriku-Haruka-Oki earthquake (M7.5) of 28 December 1994. The significance of the performance has been evaluated by a Monte Carlo study of the results since mid-1991. This shows that the Poisson model can be rejected at the 1% level, and that the results are consistent with the swarm hypothesis. The test is continuing.

1. Introduction

The current method of estimating future earthquake activity relies on the stationary Poisson model, which has been applied worldwide since the 1960's for purposes of seismic zoning. The model is thus basic to the present design of earthquake countermeasures. It is also the accepted standard against which proposed new methods of estimation are judged. If any new method turns out to be superior, countermeasures can be improved accordingly (Evison, 1982a).

The stationary Poisson model, as applied to the historical earthquake record, quantifies the probability of future earthquakes in terms of location and size, averaged over time. The probabilities do not vary with time; no estimation is provided concerning individual earthquake occurrences. To find a method that gives the variation with time, and in this respect is superior to the stationary Poisson model, is the present aim of research related to earthquake prediction, on whatever time-scale. As seismic zoning maps illustrate, the Poisson model estimates the hazard at every location in a large region. To be competitive, an earthquake prediction method should do likewise; to this end, as Suzuki (1982) pointed out, it should take account not only of successful predictions, but of failures and false alarms as well. Thus any proposed method for estimating long-term, time-varying, earthquake hazard should take the form which is standard in meteorology, i.e., long-range synoptic forecasting. The ranges of time-scales are, of course, widely different.

A proposed long-term relation between earthquake swarms and major earthquakes is the subject of a continuing study in Japan, New Zealand, and more recently, Greece. The first step in the Japan study was to define the class of multiple earthquake events and to develop recognition criteria for the three sub-classes: mainshock event, swarm and multi-

plet. It is the mainshock events that the hypothesis purports to predict, using the swarms as precursors. The multiplets, once identified, are of no further relevance. The lists of these events are an integral part of the hypothesis, and each event is tabulated in sufficient detail that every individual earthquake belonging to the event can be identified in the catalogue. Recognition criteria for mainshock events, swarms and multiplets were published by Evison (1981), together with lists of each since 1926 for the Japan region. The updated lists given in the Appendix below (Tables 1, 2, 3) take account of the JMA catalogue revision for the period 1926–1960. The lists refer mainly to the area of surveillance (Fig. 1); events are also included which contributed to recognised precursory sequences and occurred outside the area of surveillance, or before 1926. The early events are taken from Utsu (1979, 1982).

From the lists of past events, clusters of swarms can be recognized that have been precursory to clusters of mainshock events. (Here the term cluster is taken to mean one or more.) On this empirical basis, precursory relations have been found such that from the location, time of occurrence, and magnitude of a recent cluster of swarms one can estimate the same parameters of the related future mainshocks. The mainshock location is estimated from the magnitude-weighted swarm epicentres. The time between swarm and mainshock, and the mainshock magnitude, are both estimated from the swarm magnitude, M_p, defined as the average magnitude of the three largest swarm earthquakes. Precursory relations for Japan were published by Evison (1982b). Combined relations for Japan and New Zealand were subsequently obtained (Rhoades and Evison, 1993). In both regions, and also in Greece, it is apparent that precursory swarms are a characteristic of shallow seismicity in subduction zones (Evison and Rhoades, 1993, 1997).

The test reported below compares the precursory swarm hypothesis directly with the stationary Poisson model, as a
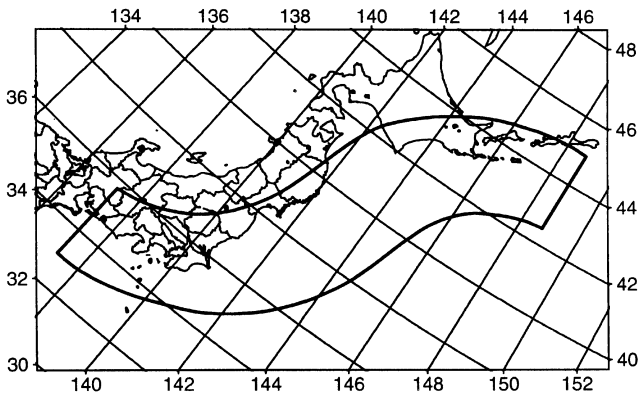
Fig. 1. Area of surveillance, shown by S-shaped outline. (Projection: Lambert polar azimuthal equivalent).

means of estimating the hazard due to all mainshocks of magnitude $M \geq 6.8$ that occur within the area of surveillance. (Magnitudes given in the present paper are JMA, unless otherwise indicated.) For the Poisson model, the average rate of occurrence of such earthquakes was taken as 9.63×10^{-7} $km^{-2}yr^{-1}$, with the Gutenberg-Richter parameter $b = 1.0$ (Rhoades and Evison, 1993). During the test period, the likelihood of the seismicity under the hypothesis, relative to that under the Poisson model, measures the performance of the hypothesis. The methodology needed for such a study has been published in mathematical form by Rhoades and Evison (1979, 1993), and will be traversed briefly in the present paper.

The test is being carried out in accordance with the International Code of Practice for Earthquake Prediction (IUGG, 1984).

2. Performance

Earthquake prediction and, more comprehensively, synoptic forecasting, concentrates the hazard towards particular points in time, location and magnitude, leaving a reduced hazard elsewhere. By contrast, the hazard under the stationary Poisson model is taken as constant in time, as uniform over the area of surveillance (at a level indicated by the historical record), and as following the Gutenberg-Richter relation for magnitude. At any point (m, t, x, y) in the magnitude-time-location space to which the forecasting hypothesis is being applied, let the hazard intensity, i.e., the local rate of earthquake occurrence, be $\lambda_1(m, t, x, y)$ under the hypothesis and $\lambda_0(m, x, y)$ under the Poisson model. Then, following Rhoades and Evison (1993), the hazard refinement factor H at that point is given by

$$H(m, t, x, y) = \frac{\lambda_1(m, t, x, y)}{\lambda_0(m, x, y)}. \tag{1}$$

When greater than unity, H has been called the probability gain (Aki, 1981).

Under the swarm hypothesis, each swarm (or swarm cluster) generates a hazard increase of the form

$$\lambda_{1j}(m, t, x, y) = f_j(m, t)g_j(x, y) \tag{2}$$

where the magnitude and time of occurrence of a related

mainshock are considered to be independent of its location. The combined hazard at any point in the magnitude-time-location space to which the swarm hypothesis is being applied is then given by

$$\lambda_1(m, t, x, y) = \sum_{j=1}^{k} \lambda_{1j}(m, t, x, y) + \theta_f(m)\lambda_0(m, x, y) \tag{3}$$

where k is the number of swarms (or swarm clusters), and $\theta_f(m)$ is the failure rate for mainshocks of magnitude m, i.e., the estimated proportion of mainshocks of magnitude m that are unrelated to swarms.

For each mainshock event that subsequently occurs, the hazard refinement factor H becomes a likelihood ratio, i.e., the likelihood of the event under the hypothesis, divided by its likelihood under the Poisson model. For the catalogue as a whole, i.e., taking account both of mainshock events that occur and of those that do not, the likelihood ratio is given by

$$LR = \prod_{i=1}^{n} H(m_i, t_i, x_i, y_i) \times \frac{P_0(\lambda_1)}{P_0(\lambda_0)} \tag{4}$$

where (m_i, t_i, x_i, y_i), $i = 1, \ldots, n$, are the points at which mainshocks have occurred, $P_0(\lambda)$ is the probability of no mainshock event occurring when the intensity of hazard is λ, and λ_1, λ_0 have been defined above. $P_0(\lambda)$ is given by

$$P_0(\lambda) = \exp\left[-\int_{m_0}^{\infty}\int_{t_a}^{t_b}\int_R\int \lambda(m, t, x, y)dydxdtdm\right] \tag{5}$$

where m_0 is the lower threshold of magnitude, t_a is the starting time, t_b is the time at which the performance is reckoned, and R is the area of surveillance. The likelihood ratio LR measures the performance of the hypothesis relative to that of the Poisson model. The algorithm for calculating the hazard refinement factor, and hence the performance factor, has been given in detail by Rhoades and Evison (1993).

A graph of the performance factor during the test is given in Fig. 2. The main performance epochs, which will be discussed in further detail below, are as follows:

1983.10.11–1994.10.04. The performance factor changed gradually, increasing from the initial value of unity to 1.55 on 1986.07.21, then decreasing to 0.95 on 1991.07.01, and increasing again to 1.25 on 1994.10.04. The performance during this epoch relates to the non-occurrence of predicted or unpredicted earthquakes; the changes of slope will be discussed in Section 5, below.

1994.10.04. The Hokkaido-Toho-Oki earthquake (M8.1), occurred on this date, and contributed an amount 2.96 to the performance factor, increasing it to 4.08.

1994.10.04–1994.12.28. The performance factor increased gradually to 4.10, due to non-occurrences.

1994.12.28. The Sanriku-Haruka-Oki earthquake (M7.5), occurred on this date, and contributed an amount 0.84 to the performance factor, decreasing it to 3.44.

1994.12.28–1998.07.31. The performance factor increased gradually to 5.04, again due to non-occurrences.

Overall, the performance shows that the mainshock activity ($M \geq 6.8$) that occurred in the area of surveillance during
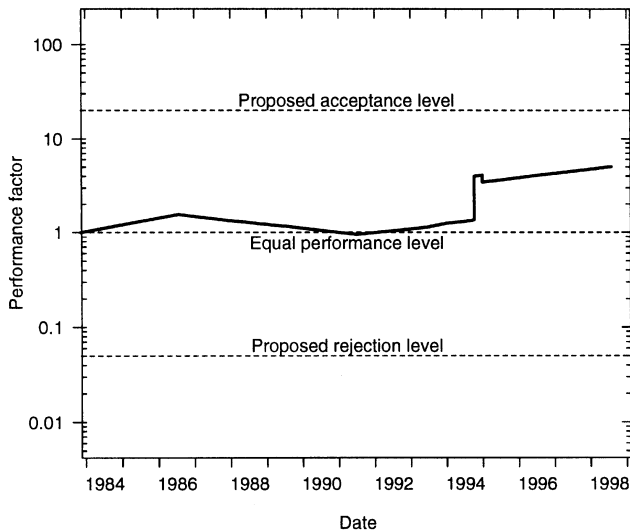
Fig. 2. Performance of precursory swarm hypothesis relative to stationary Poisson model.
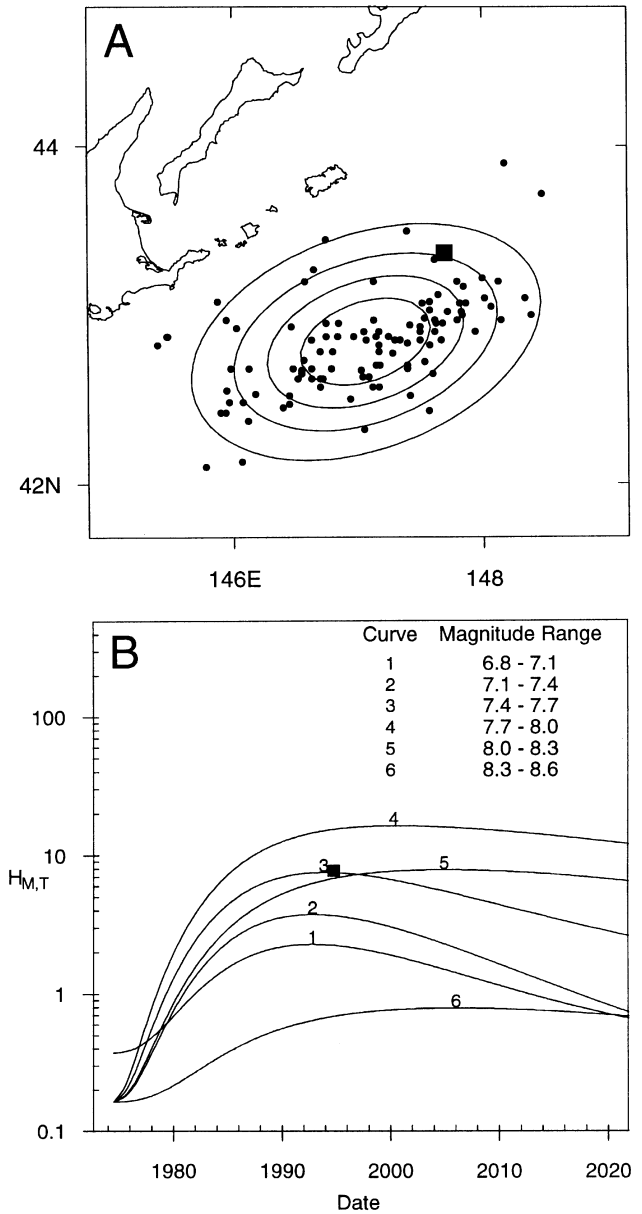

Fig. 3. Test prediction generated by precursory swarms (see text, Section 3). Data for the Hokkaido-Toho-Oki (M8.1) earthquake of 4 October, 1994, have been superimposed. (A) Location factor H_L. The value is unity at the centre of the target, and contours are shown for $H_L = 0.8, 0.6, 0.4$ and 0.2. The filled circles are the swarm epicentres, and the filled square is the mainshock epicentre. (B) Magnitude and time-of-occurrence factor $H_{M,T}$. The filled square represents the mainshock. The combined performance factor is 2.96.

the test was 5.04 times more likely under the hypothesis than under the stationary Poisson model. It was proposed at the outset that if this value rises to 20, the hypothesis as formulated will be offered for acceptance; alternatively, if the value sinks to 0.05, the hypothesis will be rejected. These values are not part of the hypothesis, but are suggested as levels at which the test might be assessed, one way or the other. In the case of rejection, either the swarm phenomenon will be abandoned as a possible precursor, or the hypothesis will be modified in the light of what has been learned during the test, and a new test will be started.

The following details of the performance graph illustrate how one evaluates the predictive role of swarms, the relation between mainshocks and swarm precursors, and the non-occurrences that are an essential feature of synoptic forecasting.

3. The Hokkaido-Toho-Oki (M8.1) Earthquake of 4 October 1994

The test results for this earthquake are shown in Fig. 3. A test prediction contains graphs of the location factor, H_L, and the magnitude and time-of-occurrence factor, $H_{M,T}$. In the notation given above, H_L is given by

$$H_L(x, y) = \frac{g_j(x, y)}{g_j(x_{j0}, y_{j0})} \qquad (6)$$

where (x_{j0}, y_{j0}) is the magnitude-weighted mean epicentre of the swarm earthquakes. $H_{M,T}$ is given by

$$H_{M,T}(m, t) = \frac{\lambda_{1j}(m, t, x_{j0}, y_{j0}) + \theta_f(m)\lambda_0(m, x_{j0}, y_{j0})}{\lambda_0(m, x_{j0}, y_{j0})}.$$
$$(7)$$

For large values of $H_{M,T}$ the performance factor is given, to a good approximation, by

$$H(m, t, x, y) \cong H_L(x, y) \times H_{M,T}(m, t). \qquad (8)$$

In all cases, the performance factor is given exactly by Eq. (1).

Details of the two swarms which generated the test prediction were published by Evison (1981, table 2, Swarms 21 and 24), and are included in the Appendix below (Table 2, Swarms 24 and 26). Given these swarms, the prediction follows directly by the published methodology, as summarised above. So also does the evaluation of the earthquake itself. The extent to which the occurrence of the Hokkaido-Toho-Oki earthquake favours the swarm hypothesis over the Poisson model is indicated by the performance factor value of 2.96.

The earlier of the two swarms that generated this predic-

tion is an event that has been much discussed in the literature, as the 1973 Nemuro-Oki (M7.4) earthquake of 17 June 1973. Mogi (1995) cites this event as a successful example of long-term prediction on the basis of seismic gap theory, but goes on to say that there have been no cases in Japan in which the place, magnitude and time of an earthquake have been successfully forecast. Kasahara and Kato (1980/81) stated that the size of the event, as inferred from magnitude, aftershock area, and tsunami source, was not large enough to fill the observed gap. The listing of the event as a precursory swarm (Evison, 1981) contributed largely to the prediction of the Hokkaido-Toho-Oki earthquake (Fig. 3).

Because of the overriding importance of objectivity in hypothesis-testing, the mainshock event is represented in Fig. 3A simply by the mainshock epicentre. Scoring for the event is thus free of subjective judgment. From a purely geophysical viewpoint, of course, it would be preferable to use the fault-break as a whole, or perhaps the aftershock area, but these are not catalogued.

4. The Sanriku-Haruka-Oki (M7.5) Earthquake of 28 December 1994

The test results for this earthquake are shown in Fig. 4. The prediction was generated by two swarms which were listed by Evison (1981, table 2, Swarms 23 and 28), and are included in the Appendix below (Table 2, Swarms 25 and 27). Updating of the prediction followed the occurrence of two further swarms in the cluster (Appendix, Table 2, Swarms 30 and 31). The mainshock is related to the swarms at a low level, and comparable examples have sometimes been classified in the literature as near-successes. The present methodology distinguishes between the questions of whether a mainshock was related to swarms, and whether it was more likely under the hypothesis or under the Poisson model. The measure of relatedness is the contribution that the first term in $\lambda_1(m, t, x, y)$ (see Eq. (3)) makes to the total. For this earthquake, the contribution is 83%, and the earthquake and swarm cluster have now been included as related events in the scoresheets (Rhoades and Evison, 1993, tables 3 and 4). On the other hand, the likelihood ratio, as given by Eq. (1), is 0.84, and this, as already indicated, is the earthquake's contribution to the performance graph (Fig. 2). Evaluation of the hypothesis as a whole is by means of the performance graph, not the scoresheet. The latter is an updated record of historical data, and is thus part of the hypothesis, and enters into the methodology as data contributing to the computation of hazard.

Swarm 31 (Appendix, Table 2) is an event that has been referred to in the literature as the Sanriku-Oki (M6.9) earthquake of 18 July, 1992. Two earthquakes of magnitude 6.9 are recorded in the catalogue close together in location and time. The event has also been interpreted as an ultra-slow earthquake, with a moment release corresponding to that of a single earthquake of magnitude M_w7.3–7.7 (Kawasaki *et al.*, 1995). It seems that the phenomenon of slow earthquakes should eventually be allowed for in the recognition criteria for mainshock events (Evison, 1981). In the meantime, if the hypothesis as formulated is sufficiently robust, its performance will not be much affected by such complications.
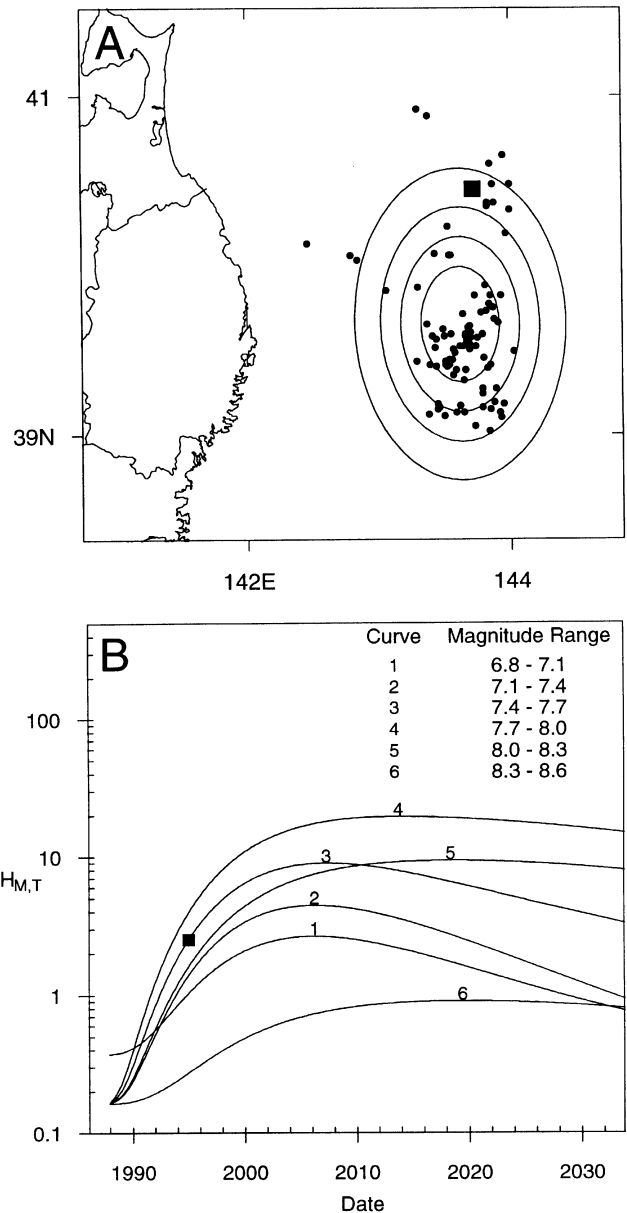

Fig. 4. Test prediction generated by precursory swarms (see text, section 4). Data for the Sanriku-Haruka-Oki (M7.5) earthquake of 28 December, 1994, have been superimposed.(A) and (B) as in Fig. 3. The combined performance factor is 0.84.

5. Non-Occurrence Factor

The performance of the hypothesis for an interval of time that passes without relevant mainshock events occurring is evaluated by the last factor in the likelihood ratio (Eq. (4)), i.e., $P_0(\lambda_1)/P_0(\lambda_0)$. This is called the non-occurrence factor. It has two components, just as occurrences of predicted and of unpredicted mainshock events are represented by the two terms in the combined hazard $\lambda_1(m, t, x, y)$ in Eq. (3).

The two non-occurrence components can be derived from Eq. (3). That for the non-occurrence of predicted mainshock events, given by $P_0(\sum_{j=1}^{k} \lambda_{1j})$, represents the cost to the performance of the increased hazard generated by swarms. Such a cost was incurred, for example, over the periods in Fig. 3B and Fig. 4B that elapsed between swarm and mainshock. The
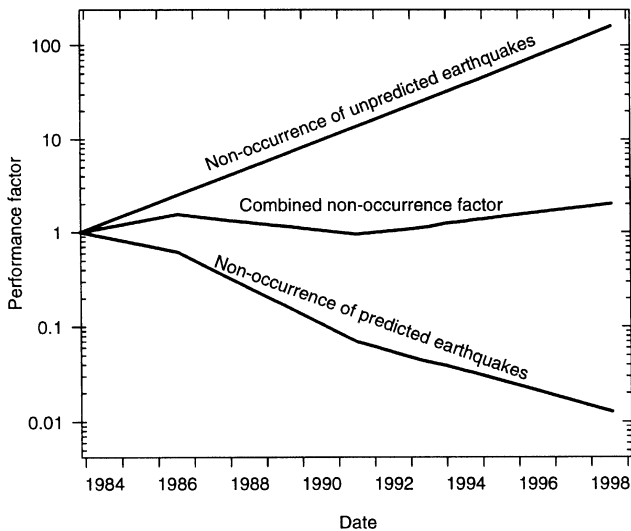
Fig. 5. Performance factors for the non-occurrence of mainshock events.

6. Significance

The ability of the test described above to distinguish, although on the basis of so few events, between the swarm hypothesis and the Poisson model, has been evaluated by Monte Carlo simulation. The real catalogue and performance graph represent but one example out of an infinity of possibilities, the distribution of which, under each model in turn, has been derived from a large number of simulated catalogues. The starting date for the simulations is taken as 1 July 1991, since there have been no subsequent changes to the algorithm, and at that time the performance factor (Fig. 2) had returned to the starting value, i.e., unity. The finishing date is taken as 31 July 1998, this being the latest date for which the catalogue data are available.

The simulations provide two large sets of performance graphs from which one calculates the two conditional probabilities known as the type I and type II errors. The type I error is the probability that the swarm hypothesis will reach the acceptance level, given that the Poisson model is correct in general; the type II error is the probability that the swarm hypothesis will not reach the acceptance level, given that it is itself correct in general. The type I error is calculated from simulations conforming to the Poisson model, and the type II error from simulations conforming to the swarm hypothesis. The simulations also indicate the extent to which the real performance differs from what would be expected under the Poisson model and swarm hypothesis, respectively.

One thousand simulated catalogues of mainshocks were generated that conformed to the Poisson model as used in the test (Rhoades and Evison, 1993). In detail, the number of mainshocks occurring over the relevant period was generated as a Poisson random variable with expected value derived from the historical rate of mainshock occurrence within the area of surveillance. The time of occurrence and location of each mainshock was chosen randomly, and the magnitudes were distributed according to the Gutenberg-Richter relation, with $M \geq 6.8$ and $b = 1$. Statistical results on the performance of the swarm hypothesis with respect to these synthetic catalogues are displayed in Fig. 6A, which also shows the real performance (as in Fig. 2). The type I error is $\ll 1\%$, since the 99th percentile of the simulations at the end of the period is much lower than the level for acceptance of the swarm hypothesis. This indicates a very small probability that the test would accept the swarm hypothesis if the Poisson model were correct. The 99th percentile is also lower than the real performance, which is thus shown to be significantly better, at the 1% level, than that expected under the Poisson model.

A second thousand simulated catalogues were generated to conform to the swarm hypothesis, with the swarm clusters in the real catalogue taken as given. The hypothesis includes provision for a proportion θ_f of mainshock events to be *unrelated* to swarm clusters, and a proportion θ_v of swarm clusters to be *related* to mainshock events (Rhoades and Evison, 1993). θ_v is adjusted for any time elapsed between swarm occurrences and the beginning of the test period. These aspects of the simulations were dealt with as follows. Mainshock events in a specially simulated series conforming to the Poisson model were randomly assigned, with probability θ_f, as unrelated to swarm clusters, and were otherwise deleted. Again, the real swarm clusters were ran-

component for the non-occurrence of unpredicted mainshock events, given by $P_0(\theta_f \lambda_0)/P_0(\lambda_0)$, represents the expected background seismicity, i.e., the expectation that unpredicted mainshock events will occur at a rate smaller than that under the Poisson model by a factor equal to the estimated failure rate. This would be the only component in the non-occurrence factor if there had been no recent swarms.

The opposing trends of the two non-occurrence components, as shown in Fig. 5, reflect the two aspects of hazard refinement under synoptic forecasting: the hazard, as already mentioned, is enhanced in certain portions of time-magnitude-location space, and diminished elsewhere. Accordingly, the non-occurrence factor is unfavourable to the performance to the extent that mainshock events are predicted, and favourable to the extent that they are not. Two sudden changes of slope can be seen in the graph for the non-occurrence of predicted earthquakes, and in the combined non-occurrence graph. The change in 1986 resulted from revisions of the catalogued magnitudes, carried out independently by the Japan Meteorological Agency and the New Zealand Seismological Observatory. Consequent upon these revisions, it became apparent that the Japan and New Zealand data could be included in the same regressions. The change of slope in 1991 resulted from algorithmic improvements in the calculation of swarm dates and of the hazard due to secondary mainshocks. (The latter are the smaller mainshocks in a cluster.)

An excessive number of predictions generated by the hypothesis would cause the bottom curve in Fig. 5 to slope more steeply downwards, and the performance curve in Fig. 2 to trend towards rejection level. For much of the test period, however, the top and bottom curves in Fig. 5 have almost counteracted each other, leaving the value of the combined non-occurrence factor at about 2; this means that the level of hazard estimated under the hypothesis has for the most part been somewhat lower than that under the stationary Poisson model.
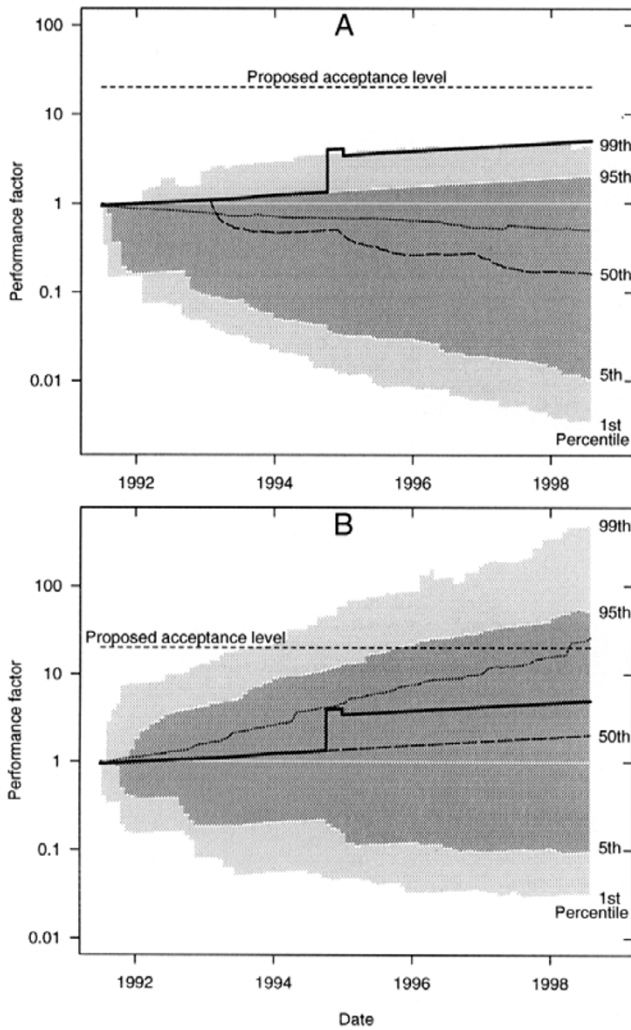
Fig. 6. Simulated performance of the precursory swarm hypothesis relative to the stationary Poisson model, for synthetic catalogues of mainshocks conforming to (A) stationary Poisson model, (B) precursory swarm hypothesis. Percentiles are indicated by shading, and the mean and median by dotted and dashed lines, respectively. The performance for the real catalogue is also shown, as in Fig. 2.
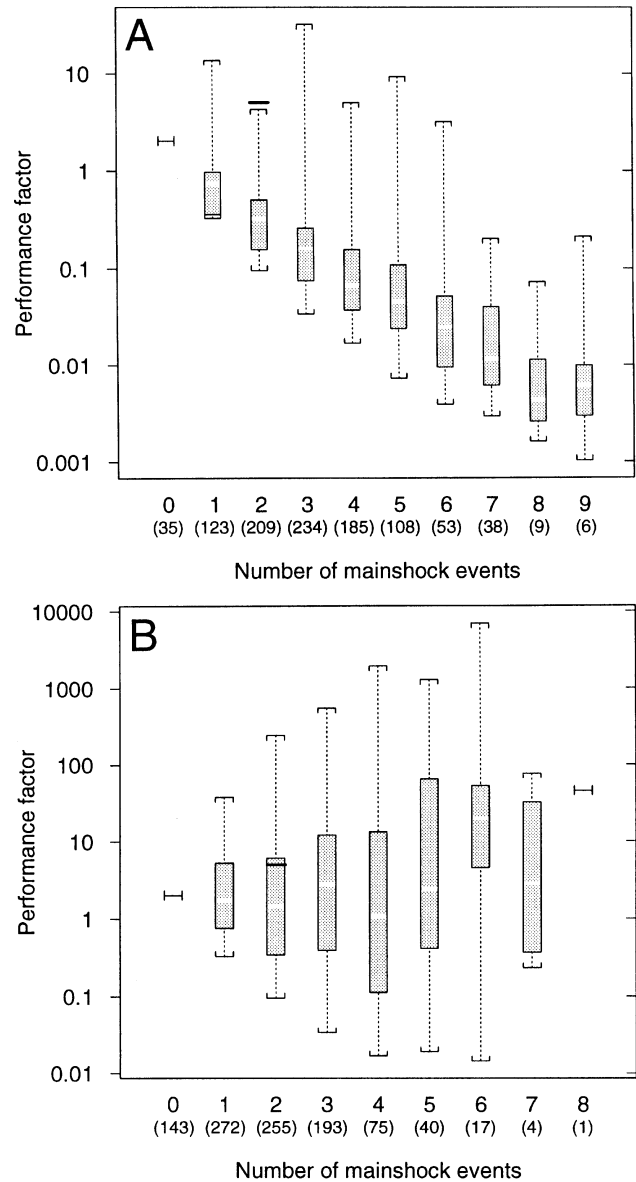

Fig. 7. Distribution of final performance factor values in Fig. 6, showing effect of number of mainshocks: (A) simulations as in Fig. 6A, (B) simulations as in Fig. 6B. In each boxplot, the box extends to the upper and lower quartiles, and the brackets to the limits of the distribution; the median is also indicated. The performance and number of earthquakes for the real catalogue are shown by a thick line. Number of simulations is shown in brackets under each number of mainshock events.

domly assigned, with probability θ_v, as related to mainshock clusters, and otherwise unrelated. The values of θ_f and θ_v were obtained from a Bayesian analysis of the scoresheet of past events (Evison and Rhoades, 1993, table 7). They are: for mainshock events, $\theta_f = 0.37$ when $6.8 \leq M \leq 7.1$, and $\theta_f = 0.16$ when $M \geq 7.2$; and for swarms or clusters of swarms, $\theta_v = 0.46$ when $5.0 \leq M_p \leq 5.6$, and $\theta_v = 0.75$ when $M_p \geq 5.7$. For the related swarm clusters, principal mainshocks were simulated randomly to conform to the swarm hypothesis, taking into account that the distribution with respect to time of occurrence is necessarily conditional on the mainshock not occurring before the start of the simulation period. Secondary mainshocks were likewise generated, with an appropriate number in each sequence, to conform to the hypothesis.

Statistical results from the simulations conforming to the swarm hypothesis are displayed in Fig. 6B. The type II error is 88%; at the end of the period, this is the percentage of simulations lying below the level for acceptance of the swarm

hypothesis; this needs to be considered, however, in the light of details not presented in Fig. 6B, and will be further discussed below. On the other hand, the performance expected under the hypothesis is not significantly different from the real performance, which lies at the 72nd percentile; any level higher than the 5th percentile would mean no significant deviation from the swarm hypothesis in the direction favouring the Poisson model.

Summarising these results, we see from the type I error that the probability was very small that the test would have accepted the swarm hypothesis if the Poisson model were correct in general. The type II error of 88% shows, on the other hand, that by the end of the period the probability was
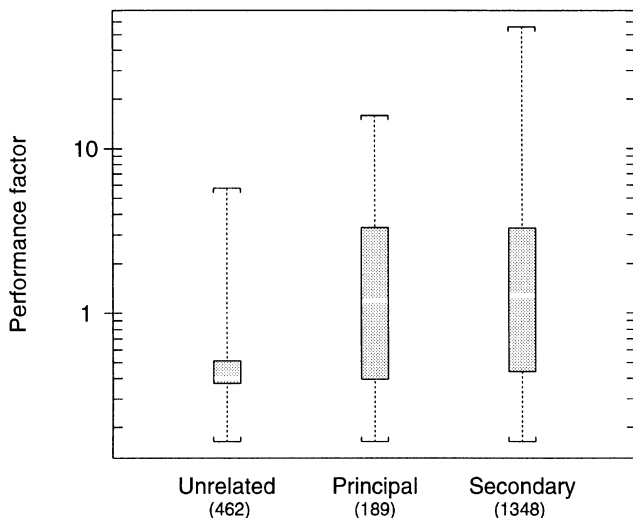
Fig. 8.   Distribution of performance factor values for each category of mainshock occurring in the 1000 simulated catalogues conforming to the swarm hypothesis (Fig. 6B). Key to boxplots as in Fig. 7. Number of mainshocks is shown in brackets under each category; the total number generated was 1999.

small that the performance factor would have reached the acceptance level proposed at the outset, if the swarm hypothesis were correct in general. The further results show that the real performance was consistent with the swarm hypothesis, and significantly better than that expected under the Poisson model.

A longer test period is needed in order to increase the number of mainshock events. The real catalogue contained two such events. The numbers in the simulated catalogues, and the corresponding performance factors, are shown by means of boxplots in Fig. 7. Under the Poisson model (Fig. 7A), 209 simulations have two mainshock events, and the real catalogue has performed better than any of them. Moreover, the performance factor for the simulations shows a downward trend as the number of mainshock events increases, and the real performance is better than all but 0.9% of them. Under the swarm hypothesis (Fig. 7B), on the other hand, a comparable number (255) of simulations have two mainshock events, and the real performance coincides with the 73rd percentile of the simulations in this category, and with the 72nd percentile of all the simulations. By comparing Figs. 7A and B, one can see that the distinction between the Poisson model and the swarm hypothesis increases generally with the number of mainshock events, up to the value 6. Thus while the real number of mainshock events during the period of the simulations was as expected under the swarm hypothesis, a higher number would give the test greater distinguishing power.

Figure 8 shows the distribution of performance factor values contributed by individual mainshocks in the simulations conforming to the swarm model, with the three categories of mainshock displayed separately. (The total number of mainshocks is, of course, the same as in Fig. 7B.) A minority of the principal and secondary mainshocks were high scoring, as already indicated by the low median value of the simulations (Fig. 6B). This feature of the test is occasioned by

the high level of historical seismicity (prior to the test) in the region of surveillance, as will be discussed in more detail below.

7.   Swarm Hypothesis Versus Poisson Model

The contrast between the swarm hypothesis and the Poisson model, as confirmed by the simulations, can be further analysed in terms of the simulated distributions of mainshock events with respect to time, magnitude and location. These distributions can also be compared with the real data, although these are so few that no statistical significance can be attached to the comparison. Figure 9A shows how the expected and actual numbers of mainshock events accumulated with time, where an expected number is given by the mean of the simulations. The expected numbers for the whole period of the simulations were 2.0 for the swarm hypothesis and 3.2 for the Poisson model. Thus the real number was the same as that expected under the swarm hypothesis. (The linearity of the Poisson graph is intrinsic to the model; the near-linearity of the swarm hypothesis graph, on the other hand, has been produced by a near-compensation of the time-variations of the hazards generated by different swarms.)

The contrast with respect to magnitude is shown in Fig. 9B, in which the ordinate is the empirical cumulative probability distribution. Starting at the large magnitude end, the Poisson graph lies above the swarm graph for $M > 8.3$, indicating a higher probability of extremely large earthquakes under the Poisson model. The swarm graph lies increasingly above the Poisson graph for $8.3 \geq M \geq 7.5$, indicating a higher probability of mainshocks in this range under the swarm model. Finally, the contribution at each magnitude is larger in the Poisson graph for $7.5 > M \geq 6.8$, indicating a higher probability in this range under the Poisson model. The real catalogue, with mainshocks at M7.5 and M8.1, thus appears to be, if anything, more compatible with the swarm hypothesis than with the Poisson model.

The contrast with respect to location is shown in Fig. 9C. Again the ordinate is the empirical cumulative probability distribution; the abscissa is the distance between the mainshock epicentre and the magnitude-weighted mean epicentre of all the earthquakes in the nearest extant swarm or swarm cluster (Rhoades and Evison, 1993). The graph shows that, for example, the proportion of mainshocks located within 100 km of the nearest swarm centre is 71% for the swarm hypothesis and 29% for the Poisson model, and both mainshocks in the real catalogue were within this distance. Thus the real catalogue appears to be more compatible with the swarm hypothesis.

Further contrasts between the Poisson model and the swarm hypothesis are evident in the simulated performance graphs. The range of the simulations at the final date, and their disposition with respect to the starting value of unity, are both very different in the two cases (Figs. 6A and B). The range between the 1st and 99th percentiles is 1000 for the Poisson and 10000 for the swarm simulations. 12% of the Poisson simulations and 66% of the swarm simulations are finally higher than unity. These differences arise as follows. The performance factor under the Poisson model is strongly affected by low-scoring mainshocks that are unrelated to swarms, while under the swarm hypothesis it is strongly af-

Fig. 10. Distribution of performance factor values for individual main-shocks in Fig. 6. Key to boxplots as in Fig. 7.
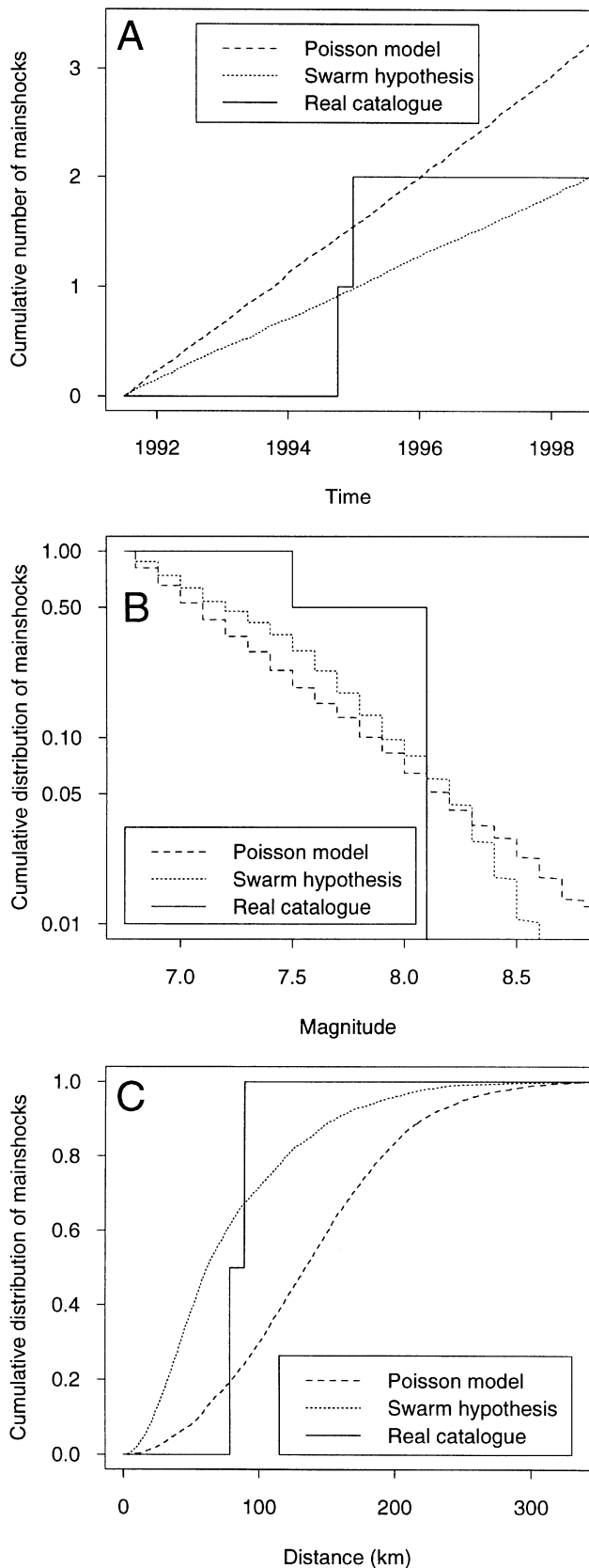

Fig. 9. Comparisons between mean of simulations under Poisson model and under swarm hypothesis, also showing real catalogue data. (A) Cumulative number of mainshocks vs time. (B) Cumulative distribution of mainshocks vs magnitude. (C) Cumulative distribution of mainshocks vs distance to nearest swarm centre.
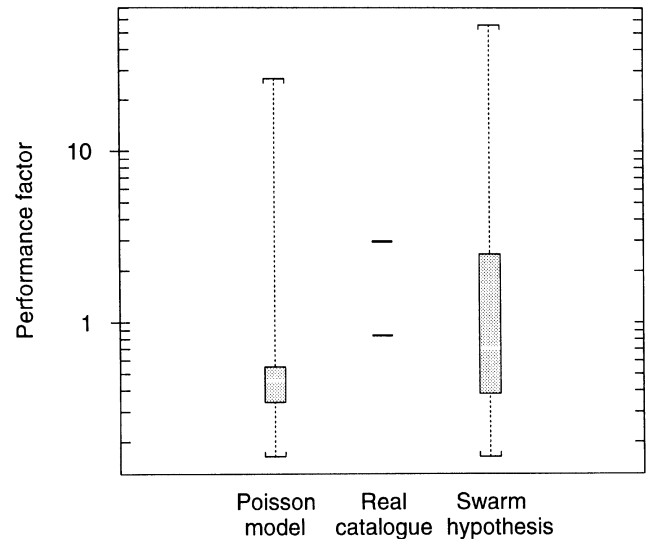
fected by high-scoring mainshocks related to swarms. But only a minority of related mainshocks are high-scoring. The mainshock scores are shown by means of boxplots in Fig. 10. The range of values, especially between the upper and lower quartiles, is much greater under the swarm hypothesis; hence the wide range in Fig. 6B. In Fig. 10 it is also clear that scores like those gained by the two real mainshocks are more compatible with the swarm hypothesis, under which they coincide with the 53rd and 79th percentiles, respectively, compared with the 89th and 98th percentiles under the Poisson model.

While demonstrating that the Poisson model and the swarm hypothesis are very different, and that the real catalogue is on the whole more likely under the swarm hypothesis, the test to date cannot be expected to have indicated acceptance of the hypothesis at the proposed level. The median value of the performance factor under the swarm hypothesis, as indicated by the 50th percentile in Fig. 6B, is about 2.0. This is 10 times higher than the median value under the Poisson model (Fig. 6A), but falls short of the proposed acceptance level by a further factor of 10.

8. Performance Versus Significance

The above analysis deals with two different ways of evaluating a prediction hypothesis: by the performance factor, or by measures of significance. In the performance factor approach, the a priori assumption is made that the hypothesis has a low probability of being superior to the current Poisson model. By proposing an acceptance level of 20, we have estimated the prior probabilities to be in the ratio 1/20 (Evison and Rhoades, 1997). The intention in choosing such a level is to guard against the swarm hypothesis being accepted (by chance) if it is not superior to the Poisson model.

In controlled experiments it is normal practice to choose the type I and type II errors in advance, and to plan the amount of experimentation with a view to achieving these levels. For any given size of experiment, the type I and type II errors vary in opposite senses. With uncontrolled experiments, as in earthquake seismology, the appropriate values of these

errors cannot be chosen in advance, because the information that they depend on comes to light only during the test. Thus one chooses an acceptance value for the performance factor. This choice, however, affects the levels of the type I and type II errors: for example, the higher the acceptance level, the lower the type I error, and the higher the type II error.

An acceptance level of 20 for the performance factor has proved more than adequate, both in the present test and in the series of tests being carried out in New Zealand, for achieving a small type I error. In the present test, as shown above, the type I error to date is $\ll 1\%$. A simulation analysis of the second test of the swarm hypothesis in New Zealand indicated a type I error of about 1% (Evison and Rhoades, 1997). Thus in both cases the chosen acceptance level has been high enough to give the tests strong protection against the swarm hypothesis gaining acceptance by chance. If the choice is too high, however, the hypothesis will for a long time fall short of gaining acceptance even if it is much superior to the Poisson model.

Evaluating the significance by means of a simulation analysis has provided a useful supplement to the present performance study and to the New Zealand test mentioned above. The latter test was inconclusive after five years, with the value of the performance factor close to unity (Evison and Rhoades, 1997). Simulations showed, nevertheless, that the real catalogue was not compatible with either the swarm hypothesis (as then formulated) or the Poisson model, and, further, that the value of the performance factor lay between what would be expected if either the hypothesis or the model were correct. This information was statistically strong enough to support abandoning the test, and starting again with a reformulated hypothesis. In the reformulation, the precursory swarm phenomenon was seen as a characteristic feature of major subduction zones; thus the New Zealand hypothesis was brought closer to the Japan version being discussed here.

In the present study the performance factor, at the value 5.04, is also inconclusive in terms of the proposed acceptance level of 20. The simulation analysis has added the information, however, that the real catalogue is inconsistent with the Poisson model at the highly significant level of 1%, and is consistent with the swarm hypothesis. As against these results, the type II error of 88% shows that, if the swarm hypothesis is correct, there has been only a 12% chance of the performance reaching acceptance level over the 7-year period of the simulations. Two circumstances affecting the performance factor may be adduced for the high type II error. The first is that, in the area of surveillance since 1926, the test period has been one of unusually infrequent mainshock events and swarms; the average rate of mainshock events ($M \geq 6.8$) fell from 0.33/yr in the pre-test period to 0.14/yr in the test period to date—a fall of 58%. The corresponding rate of swarms fell by 65%. That swarms and mainshock events have both been infrequent is favourable to the swarm hypothesis in a general way. But without mainshock events one does not have the opportunity to accumulate large performance scores, whether for successes or failures; and without swarms there is little opportunity to differentiate between the swarm hypothesis and the Poisson model. For these reasons, reaching a given level of acceptance takes longer in a period of lower seismicity.

The second circumstance affecting the performance factor is more obvious. Since the contribution of a mainshock event to the performance factor is the ratio of hazard under the swarm hypothesis to that under the Poisson model (Eq. (1)), the score for a successful prediction is inversely proportional to the historical seismicity level, other things being equal. For example, the level of historical seismicity in Japan, as adopted in the present test, is eight times higher than that adopted in parallel tests in New Zealand. Thus, if conditions were otherwise identical, the performance factor for a successful prediction would be eight times greater in New Zealand than in Japan. This effect might be offset if a wider magnitude range could be adopted in Japan, but it has been found in practice that the noise level increases with the seismicity level. The historically high and recently low seismicity levels in Japan have thus conspired to give the test reported here a very low type I error, a moderate performance level, and a high type II error. The performance result and the significance results complement one another in providing a detailed evaluation of the experiment to date.

9. Conclusion

The test results to date reject the Poisson model, at a highly significant level, in favour of the precursory swarm hypothesis. They are still inconclusive in terms of the relative performance. Continuation of the test is encouraged by the results, and more rapid progress is to be expected when earthquake activity in the Pacific subduction zone increases to a more normal level.

This systematic study of swarms and mainshock events in the main subduction region of Japan, together with a similar study in New Zealand, has provided abundant data to support a proposed long-range seismogenic model (Evison and Rhoades, 1998). An implication of the model is that in non-subduction regions the place of swarms should be taken by more protracted seismicity anomalies. This has recently been verified in the New Zealand region of continental collision (Evison and Rhoades, 1999), and suggests how the present Japan area of surveillance might be extended. A further implication of the model is that the time of onset of the seismogenic process should be preferable to the swarm-time for the purpose of estimating the time of occurrence of mainshocks. This is being investigated.

A high standard of performance needs to be demonstrated before a proposed method of synoptic earthquake forecasting can be considered for operational use. The present test will be continued accordingly.

Appendix. Tables of Multiple Events: Mainshock Events, Swarms, Multiplets

Symbols: n = number of earthquakes in the event; M_1, M_3 = largest, third-largest magnitudes in the event; l(km) = the larger of the latitude and longitude spreads of the swarm or multiplet; M_p = the average of the three largest magnitudes in the swarm or multiplet.

Table 1. Mainshock events.

No.	Time (UT)	Latitude N	Longitude E	n	$M_1 - M_3$	M_1	Locality
1	1931.03.09–1931.07.09	40°14′–41°27′	142°12′–143°16′	11	1.7	7.6	Off Tokachi
2	1932.11.26–1932.12.17	42 07 –42 25	142 28 –142 42	5	1.4	7.0	Hokkaido
3	1933.03.03–1933.10.23	38 20 –40 36	143 04 –144 55	77	1.5	8.1	Off Sanriku
4	1945.02.02–1945.06.20	40 39 –41 21	141 58 –142 35	15	1.0	7.1	Off Tokachi
5	1952.03.02–1953.08.08	41 23 –42 59	142 28 –145 43	89	1.4	8.2	Off Tokachi
6	1959.01.22–1959.01.28	37 30 –37 39	142 14 –142 46	4	1.6	6.8	Off Fukushima
7	1960.03.20–1961.06.19	39 00 –40 11	143 06 –144 04	50	1.1	7.2	Off Sanriku
8	1961.07.18–1961.07.19	29 16 –29 48	131 29 –131 46	6	1.2	6.6	Off Kyushu
9	1962.04.12–1962.04.26	37 29 –38 14	142 29 –143 20	8	1.0	6.8	Off Miyagi
10	1964.05.07–1964.05.08	40 20 –40 30	139 00 –139 31	4	1.7	6.9	Off Oga Pen.
11	1968.01.29–1968.02.07	42 40 –43 12	146 41 –147 36	18	1.2	6.9	E off Hokkaido
12	1968.05.16–1969.07.13	39 34 –41 40	142 15 –144 36	142	1.2	7.9	Off Tokachi
13	1968.06.12–1968.07.09	39 07 –39 32	142 51 –143 46	29	1.4	7.2	Off Iwate
14	1969.08.12–1971.07.10	42 33 –43 47	146 44 –148 58	123	1.5	7.8	E off Hokkaido
15	1974.05.09–1974.07.09	34 32 –34 50	138 44 –138 58	9	2.4	6.9	Izu–Oshima
16	1978.01.14–1978.01.15	34 46 –34 50	138 50 –139 15	4	1.6	7.0	Izu–Oshima
17	1978.06.12–1978.09.24	37 52 –38 28	141 57 –143 40	13	1.5	7.4	Off Miyagi
18	1980.02.23–1980.03.10	43 04 –43 27	146 33 –147 43	10	1.1	6.8	E off Hokkaido
19	1980.06.29–1980.07.07	34 51 –34 56	139 11 –139 17	7	2.1	6.7	Izu–Oshima
20	1982.03.21–1982.05.22	41 57 –42 09	142 24 –142 39	8	1.9	7.1	S off Hokkaido
21	1982.07.22–1982.08.24	36 07 –36 30	141 03 –142 03	15	1.1	7.0	Off Ibaraki
22	1983.05.26–1983.09.23	40 01 –41 30	138 45 –139 26	57	1.6	7.7	Off Tohoku
23	1994.10.04–1995.09.06	42 49 –44 05	145 34 –148 29	111	1.4	8.1	E off Hokkaido
24	1994.12.28–1995.05.25	40 02 –40 43	142 11 –144 15	26	1.0	7.5	Off Sanriku

Table 2. Swarms.

No.	Time (UT)	Latitude N	Longitude E	n	$M_1 - M_3$	l (km)	M_p	Locality
1	1901.01.14–1902.08.07	40°00′–42°48′	141°18′–144°48′	18	0.4	311	7.2	Off Tokachi
2	1923.05.26–1924.08.25	35 24 –36 30	141 00 –142 30	16	0.2	136	7.2	Off Ibaraki
3	1923.09.01–1924.01.15	34 24 –36 18	138 54 –140 36	33	0.6	211	7.5	Kanto
4	1924.12.29–1925.02.03	42 30 –43 30	147 00 –148 00	7	0.2	111	6.9	E off Hokkaido
5	1928.05.27–1929.01.11	39 08 –40 35	142 03 –143 58	37	0.5	164	6.7	Off Sanriku
6	1932.09.03–1933.01.17	40 00 –41 15	143 07 –144 39	14	0.5	139	6.6	Off Sanriku
7	1935.10.13–1935.11.19	39 57 –40 55	143 18 –144 35	17	0.6	107	6.8	Off Sanriku
8	1937.08.04–1940.02.09	35 30 –37 55	140 24 –142 57	160	0.2	268	7.4	Off Fukushima
9	1939.05.01–1941.10.08	39 54 –40 12	139 24 –140 00	9	0.6	51	6.7	Off Oga Pen.
10	1941.03.12–1941.03.19	39 20 –40 06	143 06 –144 12	6	0.1	94	6.3	Off Sanriku
11	1943.03.13–1943.05.02	35 46 –36 22	141 27 –141 55	20	0.5	67	6.4	Off Ibaraki
12	1943.06.03–1943.06.21	40 35 –41 20	142 27 –143 39	14	0.7	100	6.7	Off Sanriku
13	1948.05.12–1948.07.07	37 38 –37 56	142 21 –143 18	10	0.5	83	6.3	Off Fukushima
14	1949.10.06–1950.09.13	35 44 –36 44	139 36 –141 14	16	0.5	145	6.2	Off Ibaraki
15	1950.07.05–1950.10.26	42 06 –42 24	144 30 –145 18	5	0.5	66	5.6	Off Tokachi
16	1950.12.06–1951.07.26	29 00 –30 30	131 00 –132 00	5	0.5	167	5.7	Off Kyushu
17	1952.06.14–1952.11.01	38 54 –39 46	143 10 –143 58	29	0.1	96	6.4	Off Sanriku
18	1955.04.30–1955.06.06	39 27 –40 05	142 53 –143 53	8	0.2	85	6.0	Off Sanriku
19	1958.02.15–1958.04.11	37 54 –38 29	142 16 –143 34	13	0.5	113	6.4	Off Sanriku
20	1961.01.16–1961.02.03	35 58 –36 36	141 15 –142 31	12	0.3	113	6.6	Off Ibaraki
21	1961.02.08–1961.02.15	42 55 –43 29	147 14 –147 56	6	0.3	63	6.4	E off Hokkaido
22	1962.08.26–1962.09.07	33 46 –34 11	139 19 –139 42	12	0.5	46	5.7	Miyakejima
23	1964.11.03–1965.01.06	34 35 –34 44	138 48 –139 31	7	0.4	66	5.6	Oshima
24	1973.04.05–1973.12.02	42 06 –43 30	145 24 –148 00	60	0.9	211	7.0	E off Hokkaido
25	1974.10.10–1974.11.09	40 04 –40 55	143 18 –144 00	10	0.2	94	6.3	E off Tohoku
26	1975.05.27–1976.01.24	42 26 –43 54	147 02 –148 29	49	0.5	163	6.2	E off Hokkaido
27	1976.03.30–1976.04.19	39 30 –40 20	143 32 –144 00	9	0.1	92	5.5	E off Tohoku
28	1980.09.24–1980.09.25	35 31 –35 58	139 48 –140 13	3	0.8	50	5.6	S Kanto
29	1981.01.18–1981.12.12	38 10 –38 59	142 43 –143 48	18	0.8	94	6.6	Off Miyagi
30	1989.10.27–1990.02.10	39 01 –40 11	142 28 –144 07	42	0.8	134	6.6	Off Sanriku
31	1992.07.16–1992.08.16	39 03 –39 36	143 17 –143 57	35	0.5	61	6.7	Off Sanriku

Table 3. Multiplets.

No.	Time (UT)	Latitude N	Longitude E	n	$M_1 - M_3$	l (km)	M_p	Locality
1	1934.10.06–1934.10.06	41°30′–41°50′	143°13′–143°43′	3	0.8	41	5.8	S off Hokkaido
2	1936.12.27–1936.12.29	34 16 –34 25	138 57 –139 20	5	0.7	35	5.8	Oshima
3	1956.12.21–1956.12.27	33 40 –33 44	139 26 –139 40	5	0.6	21	5.8	Hachijojima
4	1957.11.10–1957.11.11	34 02 –34 14	139 18 –139 24	4	0.7	22	5.6	Oshima
5	1965.09.25–1965.09.25	39 30 –39 33	143 30 –143 43	3	0.1	19	5.6	Off Sanriku
6	1967.04.06–1967.04.07	34 13 –34 19	139 09 –139 10	3	0.1	11	5.2	Kozujima
7	1970.05.28–1970.06.20	40 09 –40 16	143 06 –143 19	4	0.4	18	6.0	Off Iwate
8	1971.04.06–1971.04.06	42 11 –42 17	145 24 –145 29	3	0.1	11	5.3	Off Kushiro
9	1974.01.25–1974.01.25	41 49 –41 57	144 08 –144 21	3	0.9	18	5.4	Off Tokachi
10	1975.11.11–1975.11.11	41 28 –41 37	144 14 –144 37	3	0.4	32	5.4	Off Kushiro
11	1977.12.23–1977.12.24	39 04 –39 08	143 23 –143 38	6	0.6	22	5.6	E off Tohoku
12	1982.12.27–1982.12.29	33 44 –33 52	139 22 –139 27	7	0.6	15	6.0	Miyakejima
13	1986.11.21–1986.11.22	34 33 –34 45	139 16 –139 32	3	0.9	22	5.4	Izu–Oshima
14	1992.12.28–1993.01.13	38 53 –38 56	142 33 –142 36	7	0.2	6	5.8	Off Miyagi
15	1993.03.19–1993.03.20	36 05 –36 09	141 40 –141 48	3	0.7	12	5.2	Off Ibaraki
16	1993.11.09–1993.11.09	36 13 –36 14	141 50 –141 54	3	0.2	6	5.4	Off Ibaraki
17	1997.10.09–1997.10.12	41 53 –41 57	144 58 –145 03	4	0.5	7	5.3	Off Tokachi

References

Aki, K., A probabilistic synthesis of precursory phenomena, in *Earthquake Prediction*, edited by D. W. Simpson and P. G. Richards, pp. 566–574, American Geophysical Union, Washington, D.C., 1981.

Evison, F. F., Multiple earthquake events at moderate-to-large magnitudes in Japan, *J. Phys. Earth*, **29**, 327–339, 1981.

Evison, F. F., Earthquake forecasting and countermeasures planning, *Earthquake Prediction Research*, **1**, 115–124, 1982a.

Evison, F. F., Generalised precursory swarm hypothesis, *J. Phys. Earth*, **30**, 155–170, 1982b.

Evison, F. F. and D. A. Rhoades, The precursory earthquake swarm in New Zealand: hypothesis tests, *New Zealand J. Geol. Geophys.*, **36**, 51–60, 1993.

Evison, F. F. and D. A. Rhoades, The precursory earthquake swarm in New Zealand: hypothesis tests II, *New Zealand J. Geol. Geophys.*, **40**, 537–547, 1997.

Evison, F. F. and D. A. Rhoades, Long-term seismogenic process for major earthquakes in subduction zones, *Phys. Earth Planet. Inter.*, **108**, 185–199, 1998.

Evison, F. F. and D. A. Rhoades, The precursory earthquake swarm and the inferred precursory quarm, *New Zealand J. Geol. Geophys.*, **42**, 229–236, 1999.

IUGG, Code of practice for earthquake prediction, *Int. Union Geod. Geophys. Chronicle*, **165**, 27–28, 1984.

Kasahara, K. and T. Kato, Aseismic faulting following the 1973 Nemuro-Oki earthquake, Hokkaido, Japan (a retrospective study), *Pure Appl. Geophys.*, **119**, 392–403, 1980/81.

Kawasaki, I., Y. Asai, Y. Tamura, T. Sagiya, N. Mikami, Y. Okada, M. Sakata, and M. Kasahara, The 1992 Sanriku-Oki, Japan, ultra-slow earthquake, *J. Phys. Earth*, **43**, 105–116, 1995.

Mogi, K., Earthquake prediction research in Japan, *J. Phys. Earth*, **43**, 533–561, 1995.

Rhoades, D. A. and F. F. Evison, Long-range earthquake forecasting based on a single predictor, *Geophys. J. R. Astron. Soc.*, **59**, 43–56, 1979.

Rhoades, D. A. and F. F. Evison, Long-range earthquake forecasting based on a single predictor with clustering, *Geophys. J. Int.*, **113**, 371–381, 1993.

Suzuki, Z., Earthquake prediction, *Ann. Rev. Earth Planet. Sci.*, **10**, 235–256, 1982.

Utsu, T., Seismicity of Japan from 1885 through 1925, *Bull. E.R.I.*, **54**, 253–308, 1979.

Utsu, T., Seismicity of Japan from 1885 through 1925 (Correction and supplement), *Bull. E.R.I.*, **57**, 111–117, 1982.

F. F. Evison (e-mail: Frank.Evison@vuw.ac.nz) and D. A. Rhoades (e-mail: d.rhoades@gns.cri.nz)