Earthquake Engineering, Tenth World Conference © 1992 Balkema, Rotterdam. ISBN 90 5410 060 5

A generalized semi-Markov process for spatially and temporally dependent earthquakes

Kimberly A. Lutz & Anne S. Kiremidjian
Stanford University, Calif., USA

ABSTRACT: Evidence suggests that large magnitude earthquakes on long faults exhibit both temporal and spatial dependence. Thus there is a need for a fault behavior model that incorporates these dependencies. The model presented in this paper is a generalized semi-Markov process, which is chosen for its ability to characterize the spatial dependence and represent random state transition times. The model assumes that a fault is divided into segments with well characterized behavior. It is further assumed that the stress accumulation rate is constant at all cells that comprise a characteristic fault segment. The model is applied to the northern section of the San Andreas fault. Good agreement is obtained for the expected number of events as function of magnitudes. Similarly, the probabilities of occurrences of events on this segment of the fault appear to be in close agreement with forecasts obtained by the Working Group on California Earthquake Probabilities (1990).

1. INTRODUCTION

Previous earthquake occurrence models have assumed either spatial independence or temporal independence or both. Earthquake events can be assumed to be spatially independent if the rupture zone occurs randomly along the length of the fault. A wide variety of earthquake occurrence models, including time- and slip- predictable models, assume spatial independence (Anagnos and Kiremidjian, 1984; Kiremidjian and Anagnos, 1984). Temporal dependence is the property of the interarrival time distribution being the same throughout time. Poisson models assume temporal independence (as well as spatial independence) and are concerned only with the rate of earthquakes and not with their clustering. The assumption of spatial independence may be reasonable for characteristic type earthquakes, and the assumption of temporal independence may be reasonable for small magnitude earthquakes. However, for large magnitude events (approximately moment magnitude 6.5 and above) occurring infrequently along long faults, the evidence indicates that the assumptions of temporal and spatial independence are not valid.

It has been recognized that long faults do not rupture completely during a single earthquake. This has given rise to the concept of fault segmentation, which attempts to divide a long fault into segments, each of which is capable of rupturing independently. There are studies suggesting that physical controls in the fault zone define the ends of segments, and that these segments persist through many seismic cycles (Schwartz, 1988). Thus, it can be hypothesized that earthquakes do have a spatial correlation because the rupture zones depend upon the physical controls governing the segmentation rather than being uniformly distributed over the length of the fault.

These observations highlight the need for a fault behavior model that includes both spatial and temporal dependence. In this paper we present such a model based on generalized semi-Markov process theory. The model is applied to the northern portion of the San Andreas fault.

2. GENERALIZED SEMI-MARKOV PROCESSES

The fault behavior model is specified as a generalized semi-Markov process (GSMP). A GSMP is comprised of one or more state variables that take on different values to describe the process as a function of time. A GSMP moves from state to state at random time intervals, which makes it particularly useful for modeling earthquakes. The theoretical formulation of a GSMP is treated extensively in Haas and Shedler (1985) and in Iglehart and Shedler (1983).

A GSMP is more general than a semi-Markov process. In a semi-Markov process, the probability of going from state i to state j depends only upon the states i and j; while in a GSMP, there are several different factors that affect this probability. The times between state transitions in a semi-Markov process form a holding time distribution that again depends only upon the current state and the next state. In the GSMP, not only are the states important, but the mechanism causing the state transitions also plays a role. While a GSMP retains the semi-Markovian property of moving from state to state, at random time intervals, it introduces a new level of complexity that makes it applicable to a wide range of problems, including the problem of modeling fault behavior.

A GSMP has other properties that make it useful for the problem at hand. A GSMP can have a complex state of state space copmrised of many state variables.

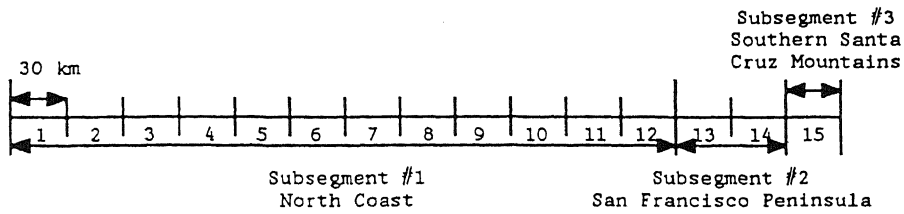

Figure 1. Model of the northern San Andreas fault with 20 km cells.

The basic physical quantity that the fault behavior model will track is the accumulated slip. Since it can be calculate from the slip rate and the elapsed time, accumulated slip is a convenient variable to describe the state of the fault. In addition, the slip released during an earthquake can be directly related to the moment release and thus the moment magnitude of the earthquake. Since the model is to have spatial dependence, it must keep track of the spatial distribution of accumulated slip. The use of a complex state space will make this possible by discretized the fault into short cells. The amount of slip accumulated on each one of these cells will then be represented by one state variable. The array of state variables defining the state of individual cells represents the state of accumulated slip on the entire fault.

3. FAULT BEHAVIOR MODEL

In order to develop the GSMP that will underlie the fault behavior model, a fault or a fault segment capable of completely rupturing in one earthquake is assumed to be composed of L subsegments with homogeneous properties (slip rate and earthquake interarrival time statistics). Each of the subsegments is further divided into cells of uniform length such that the entire fault is composed of N cells. Figure 1 shows a model of the northern San Andreas fault that was suggested by the second Working Group on California Earthquake Probabilities (1990); it has been discretized onto 30 km cells. The smallest earthquake simulated by the model corresponds to a rupture length of one cell. The fault behavior model traces through time the slip accumulated on each cell and the amount of slip release on each cell due to earthquake occurrences.

Define $\{X(t),\ t \geq 0\}$ to be a stochastic process where

$$X(t)=\{(A_1(t), B_1(t)), (A_2(t), B_2(t)), \ldots (A_N(t), B_N(t))\} \quad (1)$$

Each cell j has two state variables associated with it, $A_j(t)$ and $B_j(t)$. At time t, $A_j(t)=k$ if cell j is capable of rupturing k cells. This means that at time t, there is enough slip accumulated on cell j to cause a rupture with length of k cells. Since j is an index on the cell number, it can assume values from one to the maximum number of cells N. The number k refers to the number of cells that cell j can rupture, so it can assume values from zero (no rupture is possible) to N (the entire fault can rupture). An empirical relationship between average displacement from an earthquake and rupture length is used to relate the rupture length of a cell to the accumulated slip (Wells and Coppersmith, 1991). In this relationship it is assumed that the accumulated slip on a cell can be represented by the average displacement of an earthquake. At any time t, the value of $B_j(t)$ is either 0 or 1. This state variable is needed to strictly formulate the model as a GSMP.

All the values that the process X(t) can assume form its state space S.

$$S = \{((a_1,b_1), (a_2,b_2),(a_N,b_N)) \in (\{0,1, \ldots N\} \times \{0,1\}^N\} \quad (2)$$

Equation 2 sates that each of the a_j variables may take on values in the set $\{0,1, \ldots N\}$ and that each of the b_j variables may take on values in the set $\{0,1\}$. (The lower case aj and bj are used, rather than the upper case $A_j(t)$ and $B_j(t)$ used above, to denote specific values of the state variables, rather than their values through time.) There are no restrictions on the permissible values of the state space as defined above.

In order to determine how and when the process X(t) moves between states, it is necessary to enumerate the event set E, which contains all the different events that can occur in this process. Each event is scheduled by simulating a random number from a distribution describing the time it takes for the event to occur. A clock with this amount of time is set for the event and as the time passes this clock will always show the amount of time remaining until the event occurs. The event that triggers the state transition will be the one with the shortest amount of time on its clock, i.e. the first to occur. The event set E for the fault behavior process is

$$E = \{e_{11}, \ldots e_{N1}, e_{12}, \ldots e_{N2}, e_{13}, \ldots e_{N3}\} \quad (3)$$

For each cell j, the event e_{j1} is the event that cell j increments by one the number of cells that it is capable of rupturing. This occurs when enough time has passed that the cell has accumulated enough slip to rupture a length of fault that is longer by the length of one cell. For each cell j, the events e_{j2} and e_{j3} are events that cell j triggers and earthquake.

When the process X(t) is in a given state s (that is of necessity a subset of the entire state space S), the events that can occur and trigger a transition onto the next state must be determined. This set of events E(s) is a subset of the entire event set E defined above. The event set mapping tells how to determine which events are in E(s).

For $s = \{((a_1,b_1), (a_2,b_2),(a_N,b_N))\} \in S$:

$e_{j1} \in E(s)$ if and only if $a_j < N$

$e_{j2} \in E(s)$ if and only if $b_j = 0$ (4)

$e_{j3} \in E(s)$ if and only if $b_j = 1$

Thus for each cell j, the events e_{j1} is scheduled unless there is already enough accumulated slip on cell j to rupture the entire fault. Event e_{j2} is scheduled when $b_j = 0$, and the duplicate event e_{j3} is scheduled when $b_j = 1$.

Since the events to be scheduled in the current state have been determined, it becomes necessary to specify how to set the clock for each event. For events e_{j1} the clock is set deterministically according to an empirical equation of average displacement versus surface rupture length. Dividing the average displacement by the slip rate yields the amount of time it takes for enough slip to accumulate to rupture one cell. In order to determine the additional amount of time it takes to accumulate enough slip to rupture m cells when currently m-1 cells can rupture, the time to rupture m and m-1 cells is determined in the same manner and then the time increment is applied.

To set the clock corresponding to event e_{j2}, it is necessary to determine if cell j triggered an earthquake to cause the current state transition. If so, then the clock for e_{j2} is set by simulating a random number from the distribution D_j, which is the distribution of times between cell j triggering an earthquake. this is not the same as the distribution of interarrival times for cell j rupturing during an earthquake because cell j can break due to another cell triggering rupture. As will be explained later, the mean and standard deviation of D_j are determined by trial and error based on the estimated mean and standard deviation of the earthquake interarrival times. D_j must have only positive values so it is assumed to be lognormal, though other distributions, such as Weibull, could also be assumed.

Setting the clock corresponding to event e_{j2} when cell j did not trigger an earthquake is similar. When cell j breaks and $b_j = 0$, events e_{j3} is scheduled by simulating from D_j. If $b_j = 1$, event e_{j2} is scheduled. It becomes unnecessary to know if a given cell that ruptures triggered the transition if e_{j2} and e_{j3} are alternatively scheduled for all cells that break.

The clock speeds for the process are specified as unity. Unit clock speeds signify that the clocks run at the same rate time does; thus a clock setting equal to 5.0 years will take exactly 5 years to run to zero.

The final component that must be specified to completely describe the fault behavior process is the state transition mechanism, which is summarized by the flowchart given in Figure 2. The state transition mechanism tells how the new state s' of the [process X(t) is determined when the old state s and the trigger event set E* causing the state transition are known.

The above information completely describes the GSMP underlying the fault behavior model. To simulate the model, first set all the state variables to zero ($a_j = 0$ and $b_j = 0$ for all j). Then determine which events can occur in the present state using the event set mapping. Schedule the events using the event

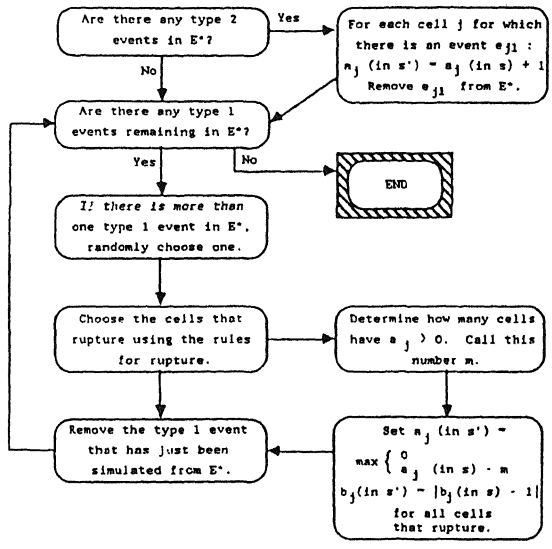

Figure 2. State transition mechanism.

scheduling mechanism. Find the trigger event by looking at the clocks for all the events that are scheduled and locating the one with the shortest time. If more than one event occurs simultaneously, the trigger event can be a set. Advance the time of the simulation by the amount of time it takes for the trigger event or trigger event set to occur. Find the new state using the state transition mechanism. If the simulation is to continue, return to the step where the events that can occur are determined from the event set mapping and go through the cycle again.

4. APPLICATION OF THE FAULT BEHAVIOR MODEL TO THE NORTHERN SAN ANDREAS FAULT

As an example of its application, the fault behavior model will be used to simulate the behavior of the northern section of the San Andreas fault, which ruptured as a whole in 1906. The fault depth will be assumed to be 20 km throughout; the assumption of

Table 1.Parameters for model application to the northern San Andreas fault

Sub-seg. #	Slip Rate (mm/yr)	Cell Size (km)	Input Mean (yr)	Input Std Dev (yr)	Output Mean (yr)	Output Std Dev (yr)	Uk Grp Mean (yr)	Uk Grp Std Dev (yr)
1	15	20	425	75	235	96	237	73
2	15	20	230	50	140	57	138	40
3	15	20	100	30	83	21	84	24
1	19	20	440	75	237	118	237	73
2	19	20	380	50	135	71	138	40
3	19	20	105	30	85	24	84	24
1	23	20	395	80	232	81	237	73
2	23	20	2000	0	101	46	138	40
3	23	20	105	30	85	24	84	24
1	15	30	435	75	241	111	237	73
2	15	30	250	75	136	65	138	40
3	15	30	86	20	85	21	84	24
1	19	30	425	75	241	118	237	73
2	19	30	600	100	120	63	138	40
3	19	30	86	20	83	23	84	24
1	23	30	400	90	234	85	237	73
2	23	30	2000	0	92	37	138	40
3	23	30	87	22	84	24	84	24

fault depth affects only the magnitude of the earthquakes that are simulated, not their occurrence in the simulation. Results are presented from eh 30 km long cells shown in Figure 2. Table 1 summarizes the parameters used in the application of the model. The second Working Group (1990) estimates the slip rate for the northern San Andreas to be 19 ± 4 mm/yr, so results will be compared using slip rates of 15, 19, and 23 mm/yr. For each subsegment, given the slip rate and the cell size, the input mean and standard deviation shown in Table 1 are the parameters of the lognormal distribution used to describe the time between earthquakes triggering on each cell of the subsegment. The output mean and standard deviation are the actual values measured when the simulation runs for a long period of time (1,000,000 years in this example). The input mean will always be the same as or larger than the output mean because cells can break when rupture spills over form an adjacent cell. For comparison, the estimates of the mean and standard deviation for each subsegment given by the second Working Group are provided. The input means and standard deviations are obtained by doing short sample runs and adjusting the input parameters to give output parameters that are reasonable.

Figure 3 shows the number of events with a given moment magnitude (M_w) or greater per year and was generated by letting the simulation run for a period of 1,000,000 years. M_w was calculated from the moment release M_0, which was calculated from the slip release, using the equation (Hanks and Kanamory, 1979:

$$M_w = (2/3) * \log (M_0) - 10.7 \qquad (5)$$

where M_0 is measured in dyne-cm. Note the relatively flat slope of the lines up to a cut-off magnitude, at which the number of earthquakes per year falls off rapidly. The largest magnitude earthquake generated for all six cases falls between $M_w = 8$ and $M_w = 8.2$, which is consistent with the assumed magnitude of the 1906 earthquake. The cut-off magnitudes are a function of the slip rate used and, to a lesser extent, the size of the cell. For a given cell size, the 15 mm/yr slip rate has the lowest cut-off magnitude and the 23 mm/yr has the highest. Since the interarrival time means remain the same when the slip rates are change, it is logical that the lowest slip rate would have the least slip accumulated and hence the smallest cut-off magnitude. The effect of the cell size is small but is most pronounced at the smallest slip rate; it shows that the longer cell size is associated with the larger cut-off magnitude.

Table 2 shows a comparison of the probabilities predicted by this model with those estimated by the second Working Group. The table contains the probability of rupturing more than 50% of each subsegment during the 30 year time period from 1990-2020 for each combination of cell size and slip rate. For subsegment 1, the Working Group estimates the rupture probability toe 0.02 while the fault behavior model found the probability to be virtually zero. this is not a large difference, especially in view of the fact that the Working Group does not consider differences of probability less than 0.10 to be meaningful. There is significant disagreement on the probability of rupturing subsegment 2, however, that reflects a difference in how the subsegment is viewed. In the Working Group

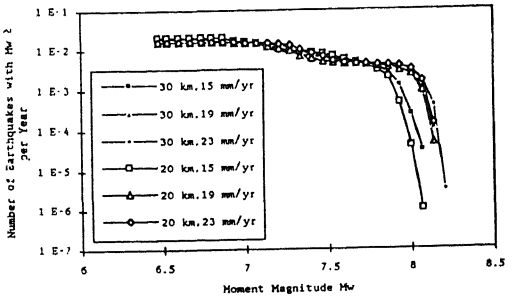

Figure 3. Cumulative number of earthquakes per year as a function of M_w.

report, this subsegment is noted to have ruptured in 1906 but not in the Loma Prieta earthquake of 1989. Based on the relatively small amount of slip released on this transition segment in 1906, the Working Group feels that subsegment has a good chance of rupturing in the next 30 years. Most of its ruptures are triggered by a rupture on an adjacent segment that spills over onto the transition subsegment. Since the fault behavior model shows a very small probability of rupture on subsegments 1 and 3, there is consequently a very small probability of rupture on subsegment 2. Both the Working Group and the fault behavior model find the probability of rupture on subsegment (which ruptured during the Loma Prieta earthquake) to be virtually zero within the next 30 years.

Figures 4-6 show the probabilities of each cell rupturing within the given time frames (50, 100, and 150 years) assuming 30 km cells and different slip rates. Note that the probability of an earthquake anywhere on the fault within 50 years is 0.07 when a slip rate of 15 mm/yr is assumed while it drops to 0.02 when a slip rate of 19 mm/yr is assumed. The reason for this is that the interarrival times are kept constant even when the slip rate is changed. When the slip rate is higher, more slip will be released in each earthquake and hence more cells will break. This means that each cell must have a higher input mean to keep the overall interarrival times the same. Because of the higher means, fewer earthquakes occur, but when they do, more cells, break. The effect of cell size can be seen by comparing Figure 5 with Figure 7, in which the cell size is 20 km. The overall probabilities are similar for the different cell sizes, but with the 20 km cells, there is a more gradual transition form the low rupture probability cells at the northern end of the fault to the high rupture probability cells at the southern end.

Table 2. Comparison of fault behavior model probabilities of rupture to Working Group estimates

Subsegment #	1	2	3
USGS Circular 1053	0.02	0.23	~0
20km cells, 15mm/yr	~0	0.033	0.0001
20km cells, 19mm/yr	~0	~0	~0
20km cells, 23mm/yr	~0	~0	~0
30km cells, 15mm/yr	~0	0.0111	~0
30km cells, 19mm/yr	~0	~0	~0
30km cells, 23mm/yr	~0	~0	~0

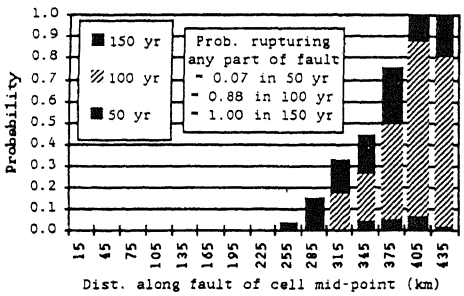

Figure 4. Probability of rupturing each 30 km cell on the northern San Andreas fault with a slip rate of 15mm/yr.

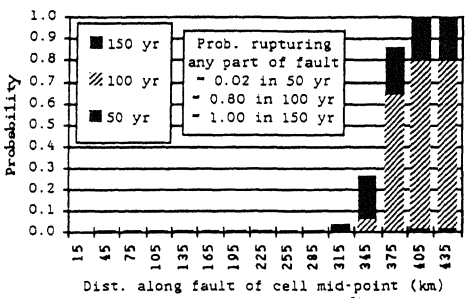

Figure 5. Probability of rupturing each 30 km cell on the northern San Andreas fault with a slip rate of 19mm/yr.

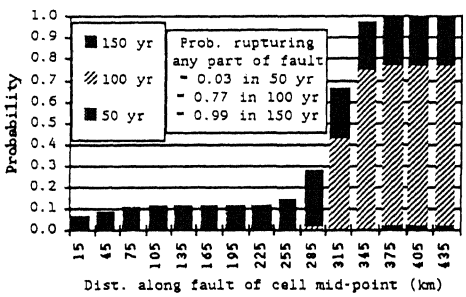

Figure 6. Probability of rupturing each 30 km cell on the northern San Andreas fault with a slip rate of 23mm/yr.

5. CONCLUSIONS AND FUTURE WORK

The presented model attempts to describe the behavior of faults that generate spatially and temporally dependent earthquakes. The model has assumed throughout that there are not stress effects; that is, the rupture of one part of the fault does not put any additional stress on the other parts of the fault. The second Working Group incorporated stress effects from the Loma Prieta earthquake into some of its probability estimates. In addition, stress effects have been included in other synthetic seismicity models

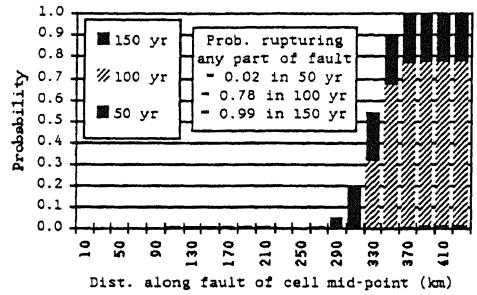

Figure 7. Probability of rupturing each 20 km cell on the northern San Andreas fault with a slip rate of 19mm/yr.

(Ward, 1991a,b) It seems reasonable to investigate further the possibility that the current model should include such stress effects.

The current model is a one-dimensional approach to the problem of earthquake modeling. It assumes that the entire depth of the fault breaks for each cell that ruptures. This is not an entirely valid assumption, as can be seen by the fact that no surface ruptures was noted for the 1989 Loma Prieta earthquake (Plafker and Galloway, 1989). A two-dimensional model could take into account such fault behavior and should be investigated.

To use the presented model to estimate site-specific hazard, a ground motion model could be used to predict the ground motion parameter of interest at the site each time an earthquake is simulated. After many simulations of the fault behavior, it would be possible to construct a graph of the values of the ground motion parameter versus the probability of exceeding the parameter, thus giving a measure of the site-specific hazard.

Finally, in order to estimate site-specific hazard in an area as tectonically complex as the San Francisco Bay Area of California, it is not sufficient to simply consider the hazard from the nearest fault. The Hayward, the Calaveras, and the northern San Andreas faults all exhibit behavior that could be simulated by this fault behavior model, and any estimate of site-specific hazard should take into account all three faults. In addition, numerous smaller magnitude earthquakes that would be damaging to structures located near the epicenter occur. Another model, perhaps a Poisson model, should be used to model the occurrence of these smaller earthquakes. Only by combining the results from both models will a picture of the total hazard at the site begin to emerge.

6. ACKNOWLEDGEMENTS

The authors would like to thank Professor Gerald Shedler who provided guidance and information about the mathematical aspects of GSMPs. The comments and discussion by J. Carl Stepp and John Schneider were very helpful and valuable in confirming the rupture mechanism. We express our gratitude for their help.

This material is based upon work supported under a National Science Foundation Graduate Fellowship. Additional support is provided under EPRI grant RP2356-91.

REFERENCES

Anagnos, T. and A.S. Kiremidjian 1984. Stochastic Time-Predictable Model for Earthquake Occurrences. *Bull. Seism. Soc. Am.*, Vol. 74, No. 6: 2593-2611.

Ellsworth, W.L., A.G. Lindh, W.H. Prescott, and D.G. Herd 1981. The 1906 San Francisco Earthquake and the Seismic Cycle. *Earthquake Prediction: An International Review.* 4:126-140.

Haas, P.J. and G.S. Shedler 1985. Regenerative Simulation Methods for Local Area Computer Networks. *IBM J. Res. Develop.* Vol. 29. No. 2: 194-205.

Hanks, T.C. and H. Kanamori 1979. A Moment Magnitude Scale. *Jour. Geophys. Res.* Vol 84, No. B5: 2348-2350.

Iglehart, D.L. and G.S. Shedler 1983. Simulation of Non-Markovian Systems. *IBM J. Res. Develop.* Vol. 27, No. 5: 472-480.

Kiremidjian, A.S. and T. Anagnos 1984. Stochastic Slip-Predictable Model for Earthquake Occurrences. *Bull. Seism. Soci. Am.* Vol. 74, No. 2: 739-755.

Plafker, G. and J.P. Galloway eds. 1989. *Lessons Learned from the Loma Prieta, California, Earthquake of October 17, 1989.* U.S. Geological Survey Circular 1045.

Schwartz, D.P. 1988. Geologic Characterization of Seismic Sources: Moving into the 1990s. *Earthquake Engineering and Soil Dynamics II - - Recent Advances in ground Motion Evaluation*: 1-42 Proceedings of the Specialty Conference on June 27-30, Park City, Utah.

Ward, S.N. 1991a. A Synthetic Seismicity Model for the Middle America Trench.*Jour. Geophys. Res.*, to appear September.

Ward, S.N. 1991b. An Application of Synthetic Seismicity Calculations in Earthquake Statistics. *Jour. Geophysi. Res.*, to appear October.

Wells, D.L. and K.J. Coppersmith February 20, 1991. Personal communication.

Working Group on California Earthquake Probabilities 1990. Probabilities of Large Earthquakes in the San Francisco Bay Region, California. *U.S. Geological Survey Circular 1053.*