The evolution of regional seismicity between large earthquakes

Geoffrey C. P. King and David D. Bowman
Laboratoire Tectonique, Institut de Physique du Globe de Paris, Paris, France

Received 16 July 2001; revised 2 July 2002; accepted 23 August 2002; accepted 23 August 2002; published 13 February 2003.

[1] We describe a simple model that links static stress (Coulomb) modeling to the regional seismicity around a major fault. Unlike conventional Coulomb stress techniques, which calculate stress changes, we model the evolution of the stress field relative to the failure stress. Background seismicity is attributed to inhomogeneities in the stress field which are created by adding a random field that creates local regions above the failure stress. The inhomogeneous field is chosen such that when these patches fail, the resulting earthquake size distribution follows a Gutenberg-Richter law. Immediately following a large event, the model produces regions of increased seismicity (aftershocks) where the overall stress field has been elevated and regions of reduced seismicity where the stress field has been reduced (stress shadows). The high stress levels in the aftershock regions decrease due to loading following the main event. Combined with the stress shadow from the main event, this results in a broad seismically quiet region of lowered stress around the epicenter. Pre-event seismicity appears as the original stress shadows finally fill as a result of loading. The increase in seismicity initially occurs several fault lengths away from the main fault and moves inward as the event approaches. As a result of this effect, the seismic moment release in the region around the future epicenter increases as the event approaches. Synthetic catalogues generated by this model are virtually indistinguishable from real earthquake sequences in California and Washington.

INDEX TERMS: 7230 Seismology: Seismicity and seismotectonics; 7260 Seismology: Theory and modeling; 7209 Seismology: Earthquake dynamics and mechanics; 7215 Seismology: Earthquake parameters; KEYWORDS: accelerated moment release, seismic cycle, earthquake prediction, regional seismicity


1. Introduction

[2] Recently two areas of study have been gaining increasing attention from the seismological community: static stress (Coulomb) interactions between earthquakes and accelerating seismic moment release before large earthquakes. In the first, it is observed that over periods of minutes to decades, future events generally occur in regions where the Coulomb stress has been enhanced by previous events. The second area of study concerns observations of increased seismic activity in a wide region around a future epicenter. The distances to which these effects occur typically extend to as much as five times the length of faulting associated with the future earthquake.

[3] In an earlier paper [Bowman and King, 2001], we showed that the dimensions of such a region can be directly related to the extent of increased Coulomb stress that must exist prior to the event, and that observations of accelerating moment release in California become much clearer when the critical region is defined using Coulomb stress. In this paper, we produce a simple physical model that explains these observations in the context of the evolution of regional seismicity associated with large earthquakes.

[4] We first establish a description of the stress field associated with major earthquakes. The description is simplified, but includes the main elements to be expected for postevent, interevent and preevent stress fields combined to give an apparent “earthquake cycle.” To this background stress field, stress heterogeneities are added, chosen to assure a standard power law, frequency-size distribution for smaller events. The distribution and moment release of these events is then examined.

[5] This model not only predicts accelerated moment release before large events, but also explains why the associated seismicity is observed at distances from, and not close to, the future epicenter. It shows that seismic activity is greatest during the aftershock sequence and before the next event and minimum in the intervening period. The variation in activity is seen to be predominantly due to changes in the “a-value” of the Gutenberg-Richter relation with associated fluctuations in the maximum magnitude of events, and not due to changes in the “b-value”.

[6] Before introducing our model of an idealised earthquake cycle and comparing it to observed data, we review previous work concerned with understanding preevent,
interevent, and postevent seismicity under the headings “Accelerating Moment Release” and “Coulomb Stress Changes.”

1.1. Accelerating Moment Release

[7] Large earthquakes are in many cases preceded by variations in regional seismicity. Some of the earliest work on this subject was conducted by Kiyoo Mogi, who demonstrated an observable increase in the level of seismicity in surrounding regions before great shallow earthquakes in Japan (the so-called “Mogi Doughnut”) [Mogi, 1969, 1981]. It was in this context that Mogi [1981] and Fedotov [1965] formulated the idea of a seismic cycle. Many authors have subsequently documented increased seismicity prior to large earthquakes. Ellsworth et al. [1981] reported an increase in the rate of earthquakes $M \geq 5$ in the years prior to the 1906 San Francisco earthquake in a broad region covering much of the San Francisco Bay area. Lindh [1990] documented similar increases in intermediate magnitude seismicity before the $1707 M = 8.3$ Kwanto, $1857 M \approx 8.1$ Fort Tejon, and 1923 $M = 8.2$ Tokyo earthquakes.

[8] In recent years, there have been efforts to describe these observed periods of enhanced seismicity using a variety of analytic functions [Raleigh et al., 1982; Sykes and Jaume´, 1990; Bufe and Varnes, 1993; Bufe et al., 1994; Sornette and Sammis, 1995; Bowman et al., 1998; Brehm and Braile, 1998; Jaume´ and Sykes, 1999]. Numerous mechanisms have been proposed to explain these observations, including heterogeneous bond-breaking models [e.g., Sahimi and Arbabi, 1996], hierarchical fiber-bundle models [e.g., Newman et al., 1994], and classical damage mechanics [e.g., Lyakhovsky et al., 2001]. Some workers have suggested that damage mechanics on a regional scale may lead to a runaway process of faulting that would be observed as an increased rate of seismicity prior to large earthquakes [e.g., Ben-Zion and Lyakhovsky, 2002]. Bufe and Varnes [1993] pointed out that a simple power law, time-to-failure equation derived from damage mechanics could be used to model the observed seismicity. They demonstrated that the seismicity before the 1989 $M = 7.0$ Loma Prieta earthquake was fit well by a relation of the form

$$e(t) = A + B(t_e - t)^m$$

where $t_e$ is the time of the large event, $B$ is negative and $m$ is usually about 0.3. $A$ is the value of $e(t)$ when $t = t_e$, i.e., the final Benioff strain up to and including the largest event. The cumulative Benioff strain at time $t$ is defined as

$$\varepsilon(t) = \sum_{i=1}^{N(t)} E_i(t)^{1/2},$$

where $E_i$ is the energy of the $i$th event and $N(t)$ is the number of events at time $t$. However, all of these models suffer from the fundamental limitation that the effects should be most prominent close to the future epicenter rather than several fault lengths away.

[9] An alternative to the above approaches has been offered by attempts to characterize large earthquakes as a critical phenomenon [Alle`gre et al., 1982; Chelidze, 1982; Rundle, 1989; Keilis-Borok, 1990; Sornette and Sornette, 1990; Alle`gre and Le Mouel, 1994; Sornette and Sammis, 1995; Saleur et al., 1996]. In these models, the large event is characterized as the end result of a process in which the stress field becomes correlated over increasingly long scale lengths. The earthquake itself is then regarded as being the equivalent of a phase change. These formulations suffer from the problem that there should be no upper limit to the size of the largest events. They also generally fail to address the observation that major events recur on major recognizable faults.

[10] Some of these problems have been addressed by modified versions of “self-organized criticality.” In so-called “intermittent criticality” [e.g., Huang et al., 1998; Sammis and Smith, 1999] (D. D. Bowman and C. G. Sammis, Intermittent criticality and the seismic cycle, submitted to Pure and Applied Geophysics, 2002, hereinafter referred to as Bowman and Sammis, submitted manuscript, 2002) the scale over which the stress field is correlated sets the size of the largest earthquake that can be expected at a particular time, but the largest events are also controlled by long-lived features of fault geometry such as segmentation. The largest event on a given segment or set of segments cannot occur until regional criticality has been achieved. This large event then destroys criticality in its associated region, creating a period of relative quiescence after which the process repeats by rebuilding correlation lengths toward criticality and the next large event. Sornette and Sammis [1995] and Sornette et al. [1996] pointed out that the power law time to failure relation in equation (1) is also to be expected if the seismic cycle is modeled as a critical phenomenon.

[11] Bowman et al. [1998] tested the predictions of this hypothesis by searching for “critical regions” before recent large earthquakes in California. They found that before all earthquakes $M \geq 6.5$ from 1950 to 1998 along the San Andreas system in California, there was a well-defined period of accelerating seismic energy release within a finite region. They further defined a “curvature parameter” that quantified the quality of the power law fit to the observed seismicity and aided in the determination of the critical region. Based on the number of events and the quality of the curve fits, it was shown that the null hypothesis that purely random clustering could generate the observed acceleration in seismicity could be rejected at greater than 95% confidence. Furthermore, they observed that the radius of the critical region found using this method scales with the magnitude of the associated large event. Subsequent studies by other workers [e.g., Brehm and Braile, 1998; Jaume´ and Sykes, 1999; Papazachos and Papazachos, 2000] have reproduced these results in the New Madrid Seismic Zone, New Zealand, and the Aegean.

1.2. Coulomb Stress Changes

[12] The observations and models described above pay little or no attention to the actual mechanism of stress transfer between events. However, many studies have demonstrated that the static stress changes from previous earthquakes have a strong influence on the location, and presumably the timing, of subsequent seismicity. Such works include the observation of stress shadows following great earthquakes [e.g., Simpson and Reasenberg, 1994; Harris and Simpson, 1996; Jaume´ and Sykes, 1996] and Coulomb stress triggering of seismicity [e.g., King et al.,
Figure 1. Schematic illustration of the stress evolution around an infinitely long fault during the earthquake cycle. The upper part of the fault is seismogenic. Below that depth, slip is aseismic. The system is driven by distant displacement (plate) boundary conditions. Orange color indicates stresses (shear stress driving fault motion) close to failure. Green to mauve indicates lower stresses. Stress values are indicated assuming an earthquake stress drop of about 100 bars. (a) Shows the stress condition immediately following the earthquake. Stress around the fault is reduced as indicated in the graph to the right to form a stress “hole” or shadow. (b) At 33% of the earthquake cycle, the upper part of the fault is locked while the lower part slips aseismically, partly filling the stress hole and reducing the stress shadow around the fault. (c) At 66% of the cycle, the stress hole continues to fill, and the stress shadow reduces. (d) Immediately prior to the next event, the upper part of the fault reaches failure stress. (e) The earthquake slip restores the stress conditions in Figure 1a, but with a further increment of slip.
1994; Stein, 1999; King and Cocco, 2001]. Common to these studies is the observation that seismicity can be strongly influenced by small changes in the static stress field. An important implication of this is that prior to an earthquake, stress must accumulate not only on the fault itself but also in a large region surrounding the fault. Consequently, defining this region has become a major objective of Coulomb stress interaction studies.

[13] The precise definition of regions of increased static stress requires modeling the contributions of events over a long period of time. Instrumental seismic data, GPS data, InSAR data and field observations of surface rupture all combine to make calculation of the stress fields of recent earthquakes routine. However, the source parameters of preinstrumental earthquakes are significantly more difficult to define. For such calculations, it is necessary to combine the historical and geological record of seismicity with the long-term loading to define regions that have been stressed by earlier events [Deng and Sykes, 1997; Stein et al., 1997; Hubert-Ferrari et al., 2000]. The difficulty of the technique is highlighted by recent studies of the Sea of Marmara. Despite the fact that the tectonic and historical information for this region is perhaps the most detailed of its kind in the world, recent models of the Marmara region following the İzmit earthquake are significantly different [Hubert-Ferrari et al., 2000; Parsons et al., 2000]. For an event such as the recent $M = 7.1$ Hector Mine Earthquake, relatively little is known of the long-term tectonic and seismic evolution of the region. Thus no attempt can be made to place it in the long-term context of the seismic and tectonic evolution of the region.

2. Modeling Stresses Through the Earthquake Cycle

[14] The loading cycle for an infinitely long vertical strike-slip fault is shown in Figure 1. This idealized fault has constant slip along strike is localized and extends through both the crust and upper mantle, which are both assumed to retain long-term strength. We note that over the timescale of the earthquake cycle, viscous relaxation in the lower crust or mantle is not required by either geological [Armijo et al., 1996, 1999; Leloup et al., 1995; Hubert-Ferrari et al., 2000, 2002] or recent GPS results [Meade et al., 2002; McClusky et al., 2001]. Thus, we use a purely elastic model to simulate the evolution of stress in the earthquake cycle.

[15] In the seismogenic zone, the evolution of the stress field is largely controlled by earthquakes. Near the surface, static friction ($\tau_s$) is greater than dynamic friction ($\tau_d$), resulting in the stick-slip cycling of stress associated with earthquakes. Beneath the seismogenic zone, static velocity-strengthening occurs [e.g., Tse and Rice, 1986], resulting in stable sliding. The boundary between stick-slip and stable sliding is not abrupt. In the “Schizosphere”, earthquake rupture that has initiated in a stick-slip zone can propagate into a stable sliding zone [Scholz, 1990]. The system is driven by the displacement boundary conditions associated with plate motions, thus subjecting the whole system to the stress necessary to drive the fault. At greater depth where plastic flow occurs, the stress required to move the fault becomes independent of normal stress. In Figure 1, an orange color indicates stress close to that necessary to cause static failure or localized plastic flow. Beneath the seismogenic zone, this remains constant. However, in the seismogenic zone, the stress field drops to the level required for dynamic rupture following an earthquake.

Figure 2. The origins and form of an inhomogeneous stress field. (a) An idealized stress concentration that reaches the failure stress. For a simple physical model, the size of the stress peak directly determines the size of the resulting earthquake when it fails. (b) When such a stressed region fails, stress is reduced along the fault. However, beyond the ends and in side lobes, stress increases. These stress concentrations, which are smaller than the original, are shown schematically. This process can explain how a stress field with inhomogeneity at all scales can be created. The largest events progressively create the stress inhomogeneities that result in smaller events. Calculating an inhomogeneous stress field in this way is impractical. (c) Although a correct inhomogeneous stress field cannot be derived, its characteristics can be defined. It must be such that when the overall stress is slightly increased, the events that result from relaxing stress peaks result in a set of events obeying appropriate scaling laws. This can be implemented in various ways (Appendix B). Peaks must be distributed in amplitude such that a larger increase in stress results in more events than a smaller one.
earthquake cycle progresses, continued stable sliding at depth, as a result of plate loading, steadily restores the region of reduced stress, Figures 1b–1d. At the end of the seismic cycle, the seismogenic zone again approaches static failure stress, and a second event occurs (Figure 1e). Figure 1 also shows the difference between the absolute stress and the failure stress (for a section perpendicular to the fault) at each step in the loading cycle.

[17] The stress fields shown in Figure 1 vary smoothly, but this is not an accurate description of the stress field in the seismogenic crust. Since faults are finite in length and active regions are subject to background seismicity, some locations must have stresses approaching the failure stress (or background seismicity would not occur), while other areas must have stresses reduced from failure stress by approximately the order of an earthquake stress drop (as a result of previous seismicity).

[18] Figures 2a and 2b give some insight into the processes that create the stress variations. Figure 2a shows a localized stress concentration prior to failure (shown as a box car). When the earthquake occurs, the stress in the failed region drops to the dynamic failure stress (as for an infinite fault), but smaller adjacent regions are subject to stress increases (again shown as having a box car form) (Figure 2b). Thus the destruction of a stress peak at one scale results in the creation of other peaks with smaller spatial dimensions. Consequently, earthquakes of any given size create stress concentrations that will be relieved by smaller future events. Using such assumptions, it is conceptually possible to calculate the synthetic stress fields necessary to derive a Gutenberg-Richter law. However, in practice, the task is intractable. Among other reasons, the precise form of the heterogeneities in the stress field will be more complex due to static interactions with other events. Dynamic processes add further complications.

[19] However, for the purposes of this paper, a complete representation of such complexity is not required. The stress peaks are assumed to determine directly the moment of the earthquake that will result when a small increase of stress (by loading) causes an event to be triggered. If the distribution of peak sizes is such that a Gutenberg-Richter distribution of event sizes results, then the true distribution of stress and the exact mechanical processes for each event become unimportant. Although simplified, this contains all of the elements of the mechanics required for later models. A one-dimensional stress distribution such as that shown in Figure 2c with the correct characteristics could be created by using a distribution such as a Cantor Dust of appropriate fractal dimension to define the lengths of flat tops of the heterogeneities. The stress values are then assigned to the peaks randomly over an appropriate range (e.g., between static and dynamic failure stresses). It is evident that if the series is sufficiently long, then even a small increase of overall stress level will bring some peaks above threshold to produce events and produce a statistically correct distribution of earthquakes.

[20] Figure 3 combines the stress evolution through the earthquake cycle of Figure 1 with an irregular stress distribution similar to that shown in Figure 2c. The stress drop for the main event is 100 bars as before, and the amplitude of the stress variations is similar. It becomes clear from this illustration why pre-event seismicity should appear at a distance from the future epicenter. In Figure 3a, immediately following the big event, no peaks in the stress field approach failure in the epicentral region. As stress levels around the past and future epicenter are reestablished, the peaks first reach failure at a distance from the future earthquake (Figures 3b and 3c). Stress levels near the future epicenter remain too low for any failure. Only immediately prior to the future event are stress levels near to the main fault restored (Figure 3e).

Figure 3. Changing stress through the earthquake cycle of an infinite length fault for inhomogeneous stress fields. (a) Immediately after the earthquake, all stress peaks near to the main event are reduced to below the failure threshold. (b) As the stress hole starts to fill, peaks first reach the threshold at large distances from the fault. (c) Progressively this pre-earthquake activity becomes more intense and approaches the future epicenter. (d) Immediately prior to the next major event, peaks on or near the main fault plane reach failure. The later modeling shows that this phase occurs very rapidly, explaining why stages shown in Figures 3b and 3c are visible for real earthquakes, but the stage in Figure 3d (classical foreshocks) [Jones and Molnar, 1979] is not unequivocally observed. (e) The stress conditions shown in Figure 3a are restored.
**Figure 4.** Stress changes for a finite length fault 80 km long and 15 km wide. The maximum stress level is the failure stress (orange), all other stress levels are lower (the color bar indicates values). The left-hand column is for the stress changes due to the main fault. For the right column, an inhomogeneous stress field is added. Black areas indicate where stress has risen above failure and earthquakes have occurred.

(a) Immediately following an earthquake. Stresses in four quadrants beyond the end of the fault and perpendicular to it (parallel and perpendicular lobes) reach the failure stress. Diagonal lobes are reduced well below failure. (f) The parallel and perpendicular lobes experience increased activity (aftershocks). Activity is suppressed in the diagonal lobes. (b) The stresses in the parallel and perpendicular lobes start to decrease in stress below the failure stress. Stress starts to rise in the diagonal lobes. (g) Aftershock activity decreases. (c) By midcycle, both the parallel and perpendicular lobes and the diagonal lobes are stressed below the failure stress. (h) The whole epicentral region is seismically quiet. (d) Toward the end of the cycle stress rises in the diagonal lobes and reduces further in the parallel and perpendicular lobes. (i) Activity becomes significant in the diagonal lobes and starts to approach the future epicenter. (e) Immediately before the next event, the diagonal lobes and the main fault have reached failure. The future aftershock zones are stress holes or shadows. (j) Activity reaches the future main event epicenter.
3. Seismicity Throughout the Earthquake Cycle for a Segmented Fault

[21] To examine seismicity through the earthquake cycle, it is necessary to calculate the stress distribution promoting activity around the main fault. We use the same assumptions as those for Coulomb modeling. The stress evolution throughout the seismic cycle is calculated by the methods described in Appendix A, and is shown in Figures 4a–4e. The stresses are calculated for the cycle of an event with magnitude 6.5 event that repeats with a stress drop of 100 bars (mean slip of 3 m, fault length of 80 km, downdip extent of 15 km). Figure 4a shows the stress distribution immediately following an earthquake. Stresses with an amplitude capable of promoting failure appear in the aftershock zones beyond the ends of the fault and perpendicular to it. Reduced stresses (or stress shadows) extend in four lobes at 45° to the main fault (the diagonal lobes). As loading continues, the stresses in the diagonal lobes gradually approach the failure stress (Figure 4d). Because the stress shadows are closest to the failure stress at their outermost extent, the diagonal lobes appear to “fill in” from their outer edges, such that the main fault segment does not approach failure stress until immediately prior to the next event (Figure 4e).

[22] Figures 4f and 4j show the same stress cycle but with an irregular stress function added. Where the stress exceeds the failure stress, the region is shaded black. Each of these regions represents an earthquake, with the size of the region related to the size of the event (see Appendix B). Several characteristics of the evolving seismicity are readily apparent in these figures. Following the main event (Figure 4f), activity concentrates in the aftershock zones and is suppressed in the diagonal lobes. The aftershock activity then diminishes (Figure 4g) until by midcycle (Figure 4h) the whole epicentral region has become seismically quiet. In the later part of the cycle, activity resumes in the diagonal lobes, migrating in toward the future epicenter as the loading continues (Figure 4i). Seismicity on the main fault segment does not resume until immediately prior to the next major event.

4. Model Results and Data

[23] The characteristics of the synthetic data can now be compared to real data in three ways. The most straightforward comparison is with observations of accelerated moment release before large events. Since many examples have now been published, the success of the model in simulating such observations is its most robust test. The model can also be compared to observations of seismicity for the Pacific Northwest region of the USA made by Bowman et al. (submitted manuscript, 2002). This concerns two events that occurred close together. The data show features that illustrate postevent, interevent and pre-event seismicity, and allows the evolution of the Gutenberg-Richter relations through time to be examined.

4.1. Accelerating Moment Release Before Large Earthquakes

[24] Bowman et al. [1998] showed that a simple search algorithm can be used to define circular regions of accelerating moment release before large earthquakes. They noted that if the region being investigated is too large, then the accelerating moment release is masked by unassociated random background seismicity, while the selection of a region that is too small excludes events that are important in defining the acceleration. It has furthermore been found that the size of the region of heightened activity found by the algorithm of Bowman et al. [1998] scales with the seismic moment of the impending large event [Bowman et al., 1998; Jaumé and Sykes, 1999]. The physical basis for these observations can be understood in terms of the model presented here.

[25] To explore these effects, we have applied the region optimization algorithm of Bowman et al. [1998] to the synthetic catalogs generated by our model. A typical result is shown in Figure 5. The model fault in this example has source parameters equivalent to a magnitude 6.5 earthquake (see caption to Figure 5). Inset in
the figure are the ratio of the residuals of a linear fit and a power law time-to-failure fit (equation (1)) as a function of radius. A low ratio of residuals indicates that the power law, time-to-failure equation provides a significantly better fit to the data than a linear fit expected for uncorrelated randomly occurring earthquakes. Our model finds a low ratio of residuals extending from a radius of about 100–280 km with the lowest values at about 180 km, consistent with the regions found with real earthquake catalogs. This can easily be understood, since the extent of the region scales with the dimensions of the stress lobes, which in turn scale with the product of the area and slip of the dislocation source (i.e., the geometric moment of the source).

The general similarity of the observed and synthetic results is also shown in Figure 6. Figures 6a–6c are the results for three earthquakes taken from Bowman et al. [1998], while Figures 6d–6f are three synthetic examples.

4.2. Evolution of Frequency-Magnitude Statistics

[26] Observations in the Pacific Northwest described by Bowman and Sammis (submitted manuscript, 2002) are summarized in Figure 7. Figure 7a shows the region where events of magnitude 5.0 and 5.4 occurred in 1990 and 1996, respectively. The regions within which accelerated moment release could be observed overlap, such that the effects associated with both events can be seen within the single circular region shown. Figure 7b shows the cumu-
relative Benioff strain release for a 15-year period including both events. The rate of seismicity is seen to increase before both events and decrease afterward, with the minimum rate occurring between the two events. Further aspects of the seismic behavior can be seen in Figure 7c. Over the 15-year period, this shows the number of events greater than magnitude 2.5, the maximum magnitude of the events and the $b$-value. These parameters show clear variations during the time period covered by the study, with the most prominent fluctuations in the total number of events and the maximum magnitude. The same effects can be seen in Figure 7d where the Gutenberg-Richter relation is plotted in a 2-year moving window over the same time period. There are clear changes in the $a$-value with associated changes in the maximum magnitude of seismicity before and after the time of the two large

**Figure 7.** Seismicity in the Pacific Northwest of the United States from 1985 to 2000 from Bowman and Sammis [2001]. (a) Map of shallow, upper plate seismicity. Stars indicate the 1990 $m_b = 5.0$ Deming earthquake and the 1996 $m_b = 5.4$ Duvall earthquake. The circle shows the region used to study the evolution of frequency-magnitude statistics in the time interval between these two events. (b) Cumulative Benioff strain in the Pacific Northwest from 1985 to 2000 for all earthquakes $m_b \geq 2.5$. Arrows indicate the times of the Deming and Duvall earthquakes. (c) Variations in the Gutenberg-Richter statistics during the seismic cycle. The data points are from a 2-year moving window. The upper curve (square symbols) is the total number of events $M \geq 2.5$ in the region at each time step. The middle curve (round symbols) is the maximum magnitude of seismicity ($M_{\text{max}}$) at each time step. The lower curve (diamond symbols) is the $b$-value at each time step calculated by the maximum-likelihood technique. The arrows indicate the time of the 1990 Deming and 1996 Duvall earthquakes. (d) Evolution of the frequency-magnitude distribution from 1985 to 2000. The statistics were calculated in a 2-year moving window. Note that both the height ($a$-value) and $M_{\text{max}}$ in the distribution slowly increase prior to 1990 and 1996. The arrows indicate the time of the 1990 Deming and 1996 Duvall earthquakes.
Figure 8. Comparison of synthetic earthquake catalog with seismicity from the Pacific Northwest. The main earthquake in the synthetic data has source parameters equivalent to a magnitude 5.5 event. For ease of interpretation, the timescale of the synthetic data is scaled such that the time of the earthquake in the main fault of the model corresponds to the origin times of the 1990 Deming and 1996 Duvall earthquakes. In this figure, the arrows correspond to the origin time of both the real and synthetic “main shocks.” For both the real and synthetic earthquake distributions, the minimum magnitude the catalog is set to $M_{\text{min}} = 2.5$. (a) Cumulative Benioff strain release for the real earthquake sequence (Data) and a synthetic catalog (Model). Note that both the sequences are similar. (b) Variations in the Gutenberg-Richter statistics for the real earthquake sequence and a synthetic catalog. The synthetic data is indicated by a solid line and labeled “Model.” The dashed line (labeled “Observed”) is the same as Figure 7c. Note that the evolution of the number of events (related to the $a$-value) shows the closest correspondence, with the main earthquakes (indicated by arrows) occurring at the times of highest overall activity. The value of $M_{\text{max}}$ tracks the variations in the number of events for both the real and synthetic data; high overall levels of seismicity correspond to a higher $M_{\text{max}}$, while lower levels of activity corresponding to a lower $M_{\text{max}}$. In contrast, the $b$-value shows no consistent variations through the cycle. (c) Evolution of the frequency-magnitude statistics for the synthetic data. The synthetic data is binned in a moving window two time steps wide. Comparison with Figure 7d shows many common characteristics. From this figure, it is clear that the slope of the distribution (the $b$-value) remains fairly constant throughout the cycle, while there is a clear increase in activity prior to the main events. While variations in the maximum magnitude of seismicity can be seen to fluctuate in the figure, it is clear that this variation is controlled by the changes in the $a$-value. The use of $M_{\text{max}}$ alone as a measure of the activity of the system is difficult due to the small number of events.
earthquakes indicated by the arrows. Systematic changes of b-value are less obvious.

[28] Figure 8 shows results from the model. To facilitate comparison with the real earthquake data described in Figure 7, the synthetic catalogs represented in Figure 8 are taken from a model region equivalent to a 120 km radius window centered on a magnitude 5.5 event. The lower graph in Figure 8a is the cumulative Benioff strain release for two earthquake cycles in the model, while the upper graph is the same quantity for the Pacific Northwest as from Figure 7b. The time and magnitude of the two large model events indicated by arrows is constrained by the model parameters, however, the time, magnitude, and location of the associated background seismicity is completely determined by the interaction of the stress field and the irregular stress function as we have described above. The model does a remarkably good job reproducing the general characteristics of the real data. Note that the seismicity in the model is not identical in successive earthquake cycles. This occurs because the background seismicity is determined by the interaction of the stress field at each time step with an irregular stress function that is regenerated at each time step. Because the irregular stress function is not constant, background events occur at different times from one cycle to another.

[29] Figure 8b shows three graphs indicating the number of events greater than magnitude 2.5, the maximum magnitude of the events and the b-value for the model, with the data from Figure 7c replotted on the same graphs. The evolution of the number of events in the model and the data show remarkable similarities, with the overall level of seismicity increasing prior to the large events (indicated by arrows). The evolution of the maximum magnitude of seismicity and the b-value are less clear. There does appear to be an increase in the maximum magnitude of background events prior to the large events in the model. However, because the largest background events at any given time step are controlled not only by the evolution of the stress field but also by its interaction with the random noise field, it is not surprising that the evolution of the maximum magnitude of seismicity is less stable than the overall level of seismicity. Finally, the b-value curve in Figure 8b shows a great deal of scatter through the cycle. There is a slight correlation between the b-value and the maximum magnitude of seismicity, with low b-values occurring at the time of the main events. However, the b-value in the model is very noisy, and may be subject to the same statistical considerations described above. Furthermore, the b-values

Figure A1. (opposite) Coulomb stress changes for an earthquake cycle. The sequence of panels in this figure starts and ends between main events. The stresses are calculated for planes parallel to the main fault and for an effective friction of zero. The upper part of the fault is assumed to be segmented and seismic (see below), while the lower part creeps aseismically. Warm colors (reds and oranges) indicate increased Coulomb stress; cold colors (mauve and blues) indicate reduced Coulomb stress. The units are arbitrary. The earthquake cycle concerns the central segment in the seismic depth range. Note the scale of the figure, which shows stress changes to distances of not much more than one fault length from the central segment. Other figures in the paper concern much larger areas. As explained in the text, all fault segments except the central segment are assumed to move either aseismically or in very small events. Since calculations to create artificial catalogs divide the cycle into 20 intervals, these small events are assumed have slips of less than 0.05 units. This requirement allows idealized earthquake cycles to be presented. The stress evolution for any geometry of segmentation with any slip history can be calculated if required. (a) No parts of the fault have moved and as a consequence there is no Coulomb stress change field. (b) All parts of the fault except the central segment have moved by 0.25 units causing the central segment to experience a stress increase. (c) Prior to the main event, all parts of the fault except the central segment have moved by 0.5 units causing the central segment to experience a stress increase. (d) An earthquake on the central segment with a slip of 1 unit causes the central segment to experience a drop in stress. (e) All parts of the fault except the central segment have moved by 0.75 units. The stress again increases on the central segment. (f) All parts of the fault have slipped by 1 unit. Stress changes are again zero as in Figure A1a.
described by the model do not agree well with those obtained for the real earthquakes in the Pacific Northwest. The combined effect of the evolution of the Gutenberg-Richter statistics in the model can be easily seen in Figure 8c, which is equivalent to Figure 7c for the Pacific Northwest. This clearly shows that the evolution of the overall level of seismicity is the dominant factor that drives the observed fluctuations in the maximum magnitude of seismicity and the $b$-value.

5. Discussion

[30] The model for the evolution of seismicity that we have described is successful in providing a physical explanation for accelerated seismic moment release before major earthquakes and the high activity following the event in the form of aftershocks. Further general features of the spatial distribution of seismicity are also reproduced by the model. For example, over much of the earthquake cycle, the epicentral region in the model is quiescent, as is commonly observed for real earthquake sequences. When activity recommences prior to a future event, it occurs around the quiescent region to form a “Mogi Doughnut” [Mogi, 1969, 1981].

[31] We have compared our model to the seismic activity associated with two magnitude $\approx$5 events in the Pacific Northwest. These events were separated by a distance of about 100 km and 6 years in time (Figure 7). The seismicity in the region described by Bowman and Sammis (submitted
manuscript, 2002) (circle in Figure 7) includes postevent, interevent and preevent activity, and can consequently be compared to our “seismic cycle” model.

[32] The overall features of the cumulative seismic moment release and the Gutenberg-Richter statistics in our model are strikingly similar to the same features observed for the real earthquake sequence in the Pacific Northwest. No two runs of the model produce precisely the same catalogs, however, the evolution of the overall statistics remain the same from model run to model run. This is consistent with our understanding of the real earthquake cycle.

[33] The model also reproduces specific features observed in the data. The results that we have obtained here suggest that the maximum magnitude of background seismicity correlates roughly with the variation in the number of events throughout the cycle. In other words, the $a$-value of the frequency-magnitude distribution varies systematically through the cycle, being at a maximum immediately before and after the main event and at a minimum between earthquakes. It is easy to understand why the model behaves in this way. Although the $b$-value does vary, it is not as clearly related to the earthquake cycle and there is no obvious reason why the model as formulated in this paper should create systematic variations.

[34] Earthquake catalogs in this model have a Gutenberg-Richter relation over the entire seismic cycle imposed a priori (Appendix B). Our goal is not to derive the origins of basic earthquake scaling laws, but rather to explore the evolution of the statistics over the earthquake cycle. Thus the basic scaling laws that we have adopted are not derived, but are reasonable assumptions. Few workers would disagree that observed scaling laws are somehow related to stress inhomogeneity, but the precise way that this relates to such issues as the geometrical complexity of faulting or the initiation, propagation, and termination of individual events is not clear. However, the concept that a small increase in stress can trigger events in any given magnitude range is likely to be shared by any more complete mechanical model. Consequently, large-scale statistical features such as changes of $a$-value or maximum earthquake magnitude over the earthquake cycle can be understood from our simple model without the added complexity of more complete and detailed mechanical models.

[35] In this paper, we present generic models. A single earthquake fault segment in our model is isolated and considered to operate independently of other seismic segments. Under these conditions, the concept of a regular, repeating cycle is realistic. This simplifying condition can be modified to create multiply faulted models where the “seismic cycle” on one fault segment interacts with and disrupts other “seismic cycles” with similarly large events on adjacent segments [Ward, 1996]. However, for this to be a worthwhile goal, it is desirable to have a suitable data set of historical seismicity that can be used for comparison. At present, such data sets are limited, which is why for this study we have chosen to compare our results to instrumental data in the Pacific Northwest of the United States. However, research efforts in California and Turkey (among other places) should provide, in the future, useful data sets over a much wider spatial, temporal, and magnitude range.

[36] Finally, the models presented here have assumed that failure is initiated by a Coulomb type of failure, with an effective friction of zero. This is a simplifying assumption that facilitates visualization of the underlying physics. However, this assumption can easily be changed to produce more spatially complex seismicity patterns. A nonzero effective friction will cause the stress lobe patterns to shift and change form slightly through seismic cycle. While the resulting stress distributions will no longer retain the simple eight lobe patterns of Figure 4, the evolution of the resulting seismicity will not be strongly different from the results presented here.

6. Conclusions

[37] In this paper, we provide a description of the earthquake process that unifies previously separate observations (accelerating moment release and stress interactions) with a physically simple description of the relations between large events and evolving seismicity. For clarity in this paper, we explain the concepts using the concept of an earthquake cycle, although we recognize that this simplifies reality.

Figure A2. (opposite) Figures A2a–A2f are equivalent to Figures A1a–A1f, except that the spatial dimension of the model now extends to a distance of more than 10 times the fault length. The simple far field lobe pattern is now visible. For Figures A2b and A2c, lobes beyond and perpendicular to the fault are negative, and those at 45° are positive. Following the earthquake (Figures A2d and A2e), the signs of the lobes reverse. Note that this very simple symmetrical pattern only occurs for an effective friction of zero in the Coulomb calculation. A nonzero value introduces asymmetry, such that the change at the time of the earthquake is not a simple reversal of sign. Figures A2a–A2f represent stress changes through the earthquake cycle, but not the actual stress levels in the system, since at no place can the model exceed the failure stress without an earthquake occurring. The fault region must have a “memory” of its previous deformation history, represented by the “Tectonic Memory Stress.” This results from the combined effects of previous earthquakes of all sizes over the history of the system. For a single earthquake cycle, the stresses relative to the failure stress are shown in Figures A2g–A2l. At no time in the cycle does stress exceed the failure stress. Different parts of the fault and surrounding regions vary from having stress levels well below failure, to just reaching failure. The main fault only experiences this condition in (Figure A2i), immediately before an earthquake. Figures A2g–A2f are derived from Figures A2a–A2f by adding the “tectonic memory stress.” The “tectonic memory stress” is the negative of the modulus of the maximum amplitude of the lobes in (Figures A2a–A2f). No function that allows stress to rise above failure is possible, thus the amplitude of “tectonic memory stress” derived in this way is a minimum. However, (much) larger amplitudes would leave the region around the fault in stress shadow everywhere and suppress aftershocks and preevent seismicity completely. For more complex systems of fault evolution where segments interact, the “tectonic memory stress” will evolve with time.
methods, however, are more general and can in due course be extended to more realistic scenarios.

[38] The model modifies Coulomb stress change models by adopting a stress field through the earthquake cycle that is related to the failure stress. In Coulomb stress models, regions of increased stress are shown as having positive stresses. The modified stress field that we have presented here accommodates the fact that, as a result of the previous tectonic history, no region ever exceeds the failure stress without an earthquake occurring. Closeness to failure is indicated by a less negative stress.

[39] Small events through the cycle are attributed to the inhomogeneous stress distribution that must exist in the seismogenic crust as a consequence of tectonic history. The nature of these inhomogeneities must be such as to create and perpetuate a stress regime that results in the Gutenberg-Richter scaling relations when stress concentrations are relieved by earthquakes. By combining suitable inhomogeneous stress fields with the overall stress evolution due to the earthquake cycle, stress peaks can be identified as earthquakes, permitting the creation of artificial catalogs of the location and magnitude of background seismicity.

[40] The catalogs, which cover regions with diameters of 10 fault lengths around the epicenters, are similar to those observed around real earthquakes, suggesting that the model that we present captures important features of the physics of regional seismicity.

Appendix A: Evolution of Stress Through the Earthquake Cycle

[41] The static stress changes due to fault motion are calculated on the assumption that the earth can be treated as an elastic half-space and faults represented by dislocation surfaces [see review by King and Cocco, 2001]. On this basis, it is straightforward to calculate the stress changes for a simplified earthquake cycle. This is shown for simple strike-slip fault in Figure A1. The stresses are the tractions experienced on surfaces parallel to the main fault and promoting the same sense of motion. The upper part of the fault can be regarded as segmented and in the seismogenic zone and the lower part is assumed to always move continuously without earthquakes. The central segment of the seismogenic zone is regarded as the “main fault segment” for the earthquake cycle. The cycle that we show starts and ends between main events. The main segment consequently moves once giving one earthquake. All other segments are considered either to creep or to move in many small events, such that they may be regarded as moving continuously. This allows us to produce a monotonically evolving stress field to describe the earthquake cycle. Except creeping faults such as the central San Andreas and perhaps for some parts of subduction zones, this is not a good assumption. Adjacent fault segments can be expected to experience substantial earthquakes. As a result, the stress evolution will not be uniform as we propose but irregular as a consequence of local events. The stress interactions between segments also mean that the concept of an earthquake cycle with regularly repeating events is an abstraction. There is, however, nothing in the computational techniques that we adopt that require the simplification. It is adopted solely to render the underlying processes easy to explain and illustrate.

[42] In Figure A1a, no part of the fault has slipped and as a consequence, there is no stress field. In Figure A1b, all parts of the fault except the central segment have slipped by 0.25 units, stressing the central segment. In Figure A1c, the stress has increased further as a consequence of 0.5 units of slip on all but the central segment. An event with 1 unit of slip then occurs on the central segment (Figure A1d) with the result that the regions previously subject to increased stress now experience a stress decrease. In Figure A1e, all parts of the fault except the central segment have displacements of 0.75 units resulting in a reduction of the amplitudes of the stress lobes. By midcycle (Figure A1f), the whole fault has slipped by 1 unit with the result that stresses are again absent (as for Figure A1a).

[43] A plan view of the stress system, but for a much larger area is shown in Figures A2a–A2f. At large distances from the fault, eight lobes are visible which change in sign during the cycle. Lobes at 45° to the fault (diagonal lobes) are positive when those, beyond the ends of and perpendicular to the fault, are negative and vice versa. Only the diagonal lobes extend to the fault. The lobe pattern is simple and remains the same throughout the cycle. If optimal Coulomb stresses for a nonzero friction were calculated, the same general features would be retained but the forms of lobes would change during the earthquake cycle.

[44] While the foregoing is a correct, simplified description of stress changes during the earthquake cycle, it is not the same as the stress system driving the fault as described in Figure 1. It differs in the following ways:

1. In the earth, the stress distant from the fault is everywhere close to the failure stress. In Figures A2a–A2f, it is everywhere close to zero.
2. Stresses on or near to the fault can reach but never exceed the failure stress. If the distant stress (green) is taken to be the failure stress then the red lobes in Figures A2a–A2f indicate regions well in excess of failure.

[45] For a real fault system, it is not sufficient to calculate only the stress changes. The fault system has been subject to a tectonic history (leaving a stress memory) that leaves an inhomogeneous prestress, which must have the following characteristics:

1. When rupture ceases on the main fault segment, stress at the fault ends and aftershock regions can be brought close to, but not beyond failure stress. If this were not true, the main fault rupture would continue or main events would occur immediately in the side lobes.
2. When the main earthquake occurs, the stress on the main fault plane can have reached, but not exceeded the failure stress. If not the event would have occurred earlier.

[46] To account for these effects, a “tectonic memory stress” must be added to it. It must have a form that annuls the positive stresses in Figures A2a–A2f. This is shown in the center of the figure and the result of adding it is shown in Figures A2g–A2l. Unlike Figures A2a–A2f, which describe only stress change, these figures can be defined as stress relative to failure stress.

Appendix B: Creating the Artificial Earthquake Catalog

[47] As noted in the text, it is straightforward to create a one-dimensional field that will simulate the observed
earthquake statistics. There are two equally valid methods of producing the required field. Either the form of the irregular stress function can be adjusted to provide the necessary distribution of stress peaks at different scales, or any irregular function can be used with the relation between the size of the stress peaks and earthquake magnitude adjusted to provide the appropriate final scaling relations.

[45] The latter approach is the easiest to implement. A simple random stress distribution is created by assigning random numbers to each element of a matrix. In this paper, we have used 300 × 300 element model space. This matrix is then added to another matrix of the same dimensions containing the stress field values appropriate for the current stage of the loading cycle. All points of the matrix that exceed the failure threshold are associated with an earthquake. The location and areas of these regions are then calculated to provide the event locations and a measure of event size. For the model realizations presented here, the size of the events typically range from 1 to 200 matrix elements in area A. These areas are then scaled such that the resulting value is, for example, equivalent to a model earthquake with a magnitude M in the range of about 1–5 (M ≈ A0.3). With the appropriate magnitude scaling coefficients, this provides a Gutenberg-Richter relation with a b-value of 1.

[49] This method has no special virtue beyond simplicity and speed. To calculate catalogs, the earthquake cycle is divided into 20 time steps (0–20). At each step, a new random stress field is created and a new earthquake distribution calculated. The same random field cannot be reused at each time step or events would occur in the same place throughout the cycle. To prevent all events from having integer times of occurrence corresponding exactly to the time step, the origin times are randomly distributed over the appropriate interval. Thus the time step zero provides origin times ranging from 0 to 0.999, and time step 20 ranges from 20.0 to 20.999. Events at exactly time 0.0 and 20.0 correspond to the beginning (an earthquake) and end (the next earthquake) of the earthquake cycle. In the figures, the time range 0.0 to 21.0 corresponds to the beginning of the earthquake cycle (0–100%).

[50] Acknowledgments. The authors acknowledge discussions with Charles Sammis and James Dolan. James Rice and Bruce Shaw provided helpful reviews. This work was supported by a Chateaubriand Postdoctoral Fellowship sponsored by the French Embassy to the United States, NSF grant EAR-0107129, the EC Environment Program [PRESAP], and l’ACI de INSU (INSU contribution 292). IPGP contribution 1829. SCEC contribution 690.

References


---

D. D. Bowman, Department of Geology, California State University, Fullerton, Fullerton, CA 92834-6850, USA. (dbowman@fullerton.edu)

G. C. P. King, Laboratoire Tectonique, Institut de Physique du Globe de Paris, 4, place Jussieu, 75252, Paris Cedex 05, France. (king@ipgp.jussieu.fr)