Postseismic stress transfer explains time clustering of large earthquakes in Mongolia

Jean Chéry *, Sébastien Carretier, Jean-François Ritz

Laboratoire de Géophysique, Tectonique et Sédimentologie, CNRS and Université de Montpellier II, 34095 Montpellier, France

Received 13 July 2001; received in revised form 20 September 2001; accepted 12 October 2001

Abstract

Three M > 8 earthquakes have occurred in Mongolia during a 52-year period (1905–1957). Since these earthquakes were well separated in space (400 km), the coseismic stress change is far too low (0.001 bar) to explain mutual earthquake triggering. By contrast, postseismic relaxation gradually causes a significant stress change (0.1–0.9 bar) over large distances. Using a spring-slider model to simulate earthquake interaction, we find that viscoelastic stress transfer may be responsible for the earthquake time clustering observed in active tectonic areas. Therefore, revealing postseismic strain by satellite geodesy and modeling earthquake clusters could improve our understanding of earthquake occurrence, especially in zones where large earthquakes have already struck in past decades. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: earthquake cluster; fault; Mongolia; stress transfer; postseismic

1. Earthquake time clusters

The occurrence of several earthquakes after a long period of quiescence in a tectonic province, called earthquake clustering, is a perturbing behavior of the lithosphere. It means that a large earthquake occurring in an apparently dormant zone could be a warning signal for a subsequent one in the years or decades to follow. These clustering phenomena have been documented since the 1980s for rapid faults on certain plate boundaries. For example, paleoseismic investigations on the Southern San Andreas Fault produced evidence of clusters of three to four large earthquakes during 200-year periods followed by a long quiescence of 400 years [1]. In addition, historical sequences on the North Anatolian Fault reveal that this major fault was relatively quiescent for centuries [2,3] prior to the destructive sequence of 10 M 6–7 earthquakes starting in 1939 and ending with the Duzce earthquake in 1999 [4]. Due to the low occurrence of earthquakes, fault activity on plate interiors where the lithosphere is slowly deforming is more difficult to document. In such cases, fault motion with long-term slip rates of 0.01–1 mm/year may be difficult to prove since no historical record exists. However, the occurrence of earthquake clusters is documented in numerous areas such as the eastern Mediterranean [3,5,6], Iran [7] or in the Great Basin in the western USA [8]. Understanding why earthquakes occur-
cur in clusters, rather than in an evenly distributed manner, is a key point in assessing earthquake probability on faults. We propose that this behavior may be physically described using the idea that postseismic stress relaxation of the lithosphere after one large event may increase the crustal stress up to the failure threshold on other faults.

We test our model using a sequence of large earthquakes that occurred in Mongolia during the 20th century (Fig. 1). This sequence is due to crustal shortening between India and Eurasia (e.g. [9]). Under this slow compression, three $M > 8$ earthquakes have occurred during the last century: Bolnay (1905, July 23), Fuyun (1931, August 10) and Gobi–Altay (1957, December 4) [10,11]. An $M$ 7.8 earthquake also occurred 14 days prior to the Bolnay earthquake (Tsetserleg, 1905, July 9). All these earthquakes were located on very large faults showing a clear geomorphological expression, suggesting the occurrence of repeated earthquakes. Given that historical records of earthquakes are quite scarce in Mongolia, and that paleoseismic data are preliminary, it is still unknown whether other earthquake clusters have occurred in the past. However, recurrence time of large Mongolian earthquakes is assessed by direct field measurements. Preliminary paleoseismic studies on the Bogd Fault along which the 1957 Gobi–Altay $M$ 8 earthquake occurred show large recurrence times of “thousands of years” [12]. In addition, 100 m offsets of 80,000-yr-old alluvial fans along this fault lead to a long-term slip rate of 1.2 mm/yr [13] in accordance with a time return of 4000 years for 5 m of average slip for one earthquake. If we assume for the purpose of this argument that the mean recurrence times on Bolnay, Fuyun, and Gobi–Altay faults are in the order of 4000 years, then there is only a small probability that each of these faults undergoes a

![Fig. 1. Simplified schematic tectonic setting of Mongolia. Thick lines represent the surface fault rupture during the 20th century. Thin lines represent the Quaternary faults [51]. Solid arrows indicate the approximate shortening direction.](image)
large rupture in a 52-yr interval. Given that this did occur, we postulate that a causal relation exists between these earthquakes, and investigate physical reasons that may cause such behavior.

2. Earthquake recurrence model and coseismic stress triggers

Our current understanding of earthquake occurrence is based on the simple idea that a fault plane is progressively loaded during slow plate motion. As a first approximation, we consider that both the interseismic stress rate $\sigma$ between two earthquakes and plate velocity is constant far from active zones. When the shear stress acting on the fault reaches a static stress $\sigma_{\text{yield}}$, slip instability occurs and elastic motion of the crust releases the stress on the fault plane until the static stress $\sigma_{\text{base}}$ is reached. In this simple model, the recurrence time $T$ of the fault is given by $\frac{\sigma_{\text{yield}}}{c \sigma_{\text{base}}}$ where $\sigma_{\text{yield}}$ and $\sigma_{\text{base}}$ do not evolve with time. Variation in one of these stresses (Fig. 2a) leads to an irregular earthquake occurrence [14].

Due to the elastic behavior of the lithosphere over short time periods, a permanent strain and stress change is caused on surrounding faults after the occurrence of an earthquake (e.g. [15]). Such a distant trigger may increase the slow secular loading rate that would advance the time of an earthquake [16]. Alternatively, the fault plane may be unloaded which would delay the occurrence of the upcoming earthquake [17,18]. Because the limit stress on a fault is generally thought to be controlled by the fault friction $\mu$, the static stress change due to the coseismic fault motion of the distant earthquake should depend on the shear stress change $\Delta \sigma_t$, the normal stress change $\Delta \sigma_n$, and the fluid pressure change $\Delta P$ within the fault. The resulting stress change $\Delta \text{CFS}^{\text{co}}$ called the Coulomb failure stress (CFS) change, is given by:

$$\Delta \text{CFS}^{\text{co}} = \Delta \sigma_t - \mu \cdot (\Delta \sigma_n - \Delta P) \tag{1}$$

CFS changes as low as 0.1 bar have been invoked in many cases to be the cause of a significant seismicity increase in zones where the stress approaches to the Coulomb limit (e.g. [15,16,19]). Alternatively, a stress decrease with respect to the CFS change often correlates with a microseismicity decay or even suppression of large earthquakes [17,20]. The effect of CFS change on earthquake time occurrence can be evaluated using the concept of ‘clock advance’ $\Delta t$ [21–23]. This time difference is defined as the time by which an earthquake is advanced (or delayed if negative) due to a distant stress triggering. If the secular stress loading rate is linear, $\Delta t$ is given by $\frac{\Delta \text{CFS}^{\text{co}}}{\sigma}$ (see Fig. 2b). Since $\sigma$ can be evaluated as $\Delta \text{CFS}^{\text{co}}/T$ (see Fig. 2a), the clock advance depends on the static stress drop, the CFS change, and the
recurrence time of the fault:

$$\Delta t = \frac{\Delta \text{CFS}^{\text{co}}}{\Delta \tau^{\text{co}}} - T$$

Modeling earthquake cycles on a group of faults therefore requires a knowledge of the ratio between the distant stress change $\Delta \text{CFS}^{\text{co}}$ and the local stress drop $\Delta \tau^{\text{co}}$. Static stress drop during a large earthquake is generally estimated with a dislocation model and is expressed as:

$$\Delta \tau^{\text{co}} = C G \frac{\bar{u}}{W}$$

where $C$ is a geometrical constant close to 1 (see [24]), $G$ the elastic shear modulus, $\bar{u}$ the average displacement during an earthquake and $W$ the fault width. For values of 5 m and 15 km for $\bar{u}$ and $W$ respectively, a stress drop of 100 bar is found. Such a stress drop corresponds to the upper bound given by moment magnitude analysis [25] and is compatible with estimates deduced from radiated energy during earthquakes [26].

CFS change cannot be estimated as simply as for the local stress drop $\Delta \tau^{\text{co}}$ because the stress vector variation on a triggered fault is dependent on the relative position and orientation of the two faults. Using the geometry of Fuyun and Gobi–Altay faults, we compute the coseismic CFS change induced by the 1931 Fuyun earthquake on the Gobi–Altay Fault (Fig. 2c). We assume that the 1931 Fuyun earthquake can be modeled as a vertical, rectangular strike-slip fault ($L = 180$ km, $W = 20$ km, $\bar{u} = 10$ m) embedded in an infinite half-space. We use the formulation of Okada [27] in order to compute the stress perturbation $\sigma_n$ and $\sigma_t$ on the Gobi–Altay Fault. We compute the coseismic stress change $\Delta \text{CFS}^{\text{co}}$ assuming that the Gobi–Altay Fault friction $\mu$ is equal to 0.3, which is an intermediate value between a weak fault such as the San Andreas Fault ($\mu \approx 0.1$) and a fault as strong as the crust ($\mu$ in the range of 0.6–0.8). Our calculation yields a positive $\Delta \text{CFS}^{\text{co}}$ of $\approx 0.001$ bar (Fig. 2c). Under these circumstances, it is unlikely that a large earthquake on the Fuyun Fault would affect, either positively or negatively, the earthquake occurrence on the Gobi–Altay Fault. Computing similar CFS values between the Bolnay, Fuyun and Gobi–Altay faults using the same principle as above yields CFS values consistently smaller than 0.05 bar. This indicates that coseismic triggering between these distant faults is highly unlikely. Because static stress is clearly too small to cause distant triggering, seismic waves interacting with crustal fluids have been cited to explain anomalous distant microseismicity, as it happened in the days following the Landers earthquake [28,29]. Therefore, this mechanism may have been efficient for the triggering of the 1905 Bolnay earthquake 14 days after the Tsetserleg earthquake. However, the Fuyun earthquake and the Gobi–Altay earthquake occurred decades after the 1905 pair. For this reason we explore how postseismic stress relaxation slowly modifies the local stress drop $\Delta \tau$ and the distant stress change $\Delta \text{CFS}$ during the seismic cycle.

3. Postseismic stress relaxation

Postseismic motion at the surface occurs in response to the coseismic dislocation and seems to involve deep changes in the lower crust or in the mantle [30–32]. The progressive decay of postseismic motion can be characterized by a relaxation time $R$ ranging from a few days to several decades such as after the 1906 San Francisco earthquake [33]. Detailed postseismic motion distribution at depth is difficult to assess from surface measurement only, and can be attributed either to fault afterslip [34] or to a viscoelastic crustal relaxation [35]. In both cases, this induces a slow ‘reloading’ of the seismogenic layer [36] that brings the fault plane stress closer to the yield limit than in a model where relaxation does not occur [37]. As a consequence, the coseismic stress drop $\Delta \tau^{\text{co}}$ evolves towards a relaxed stress drop $\Delta \tau^{\text{rel}}$ as depicted in Fig. 3a. Because postseismic relaxation occurs aseismically, the associated local fault stress change cannot be estimated by seismological means. This stress change may, however, be inferred if a postseismic slip model is assumed for the deep part of the fault. In this case, the lower crustal stress relaxes to zero in this layer during the postseismic phase, meaning that the stress singularity that occurred at depth $W$ during the co-
seismic stage progressively vanishes. The corresponding relaxed stress after the postseismic relaxation becomes controlled by a dislocation of length $L$ with an infinite width. This new fault scaling leads to values of 5-20 bar for $L$ values ranging from 300 km to 75 km when replacing $L$ by $W$ in Eq. 3. In this simple model, the ratio between the relaxed local stress $\Delta \tau_{rel}$ and the coseismic local stress $\Delta \tau_{co}$ is therefore given by $W/L$. Away from the fault, the local stress change promotes a larger strain distribution as demonstrated by the remote geodetic motions detected after the Alaska earthquake in 1964 [38]. As a result, the net effect of the CFS change on a remote fault superimposes a slowly evolving postseismic change $\Delta \text{CFS}_{\text{post}}$ (Fig. 3b) on a sudden coseismic change $\Delta \text{CFS}_{\text{co}}$. The sum effects result in a relaxed stress change $\Delta \text{CFS}_{\text{rel}}$ after complete lower crustal relaxation.

Using the same approach used to evaluate the coseismic CFS change induced by 1931 Fuyun earthquake on the Gobi–Altay Fault, we now compute the relaxed CFS change. We assume that a complete postseismic relaxation after an earthquake may be simply approximated by a deep postseismic slip [39] as done, for example, for the postseismic motion of the 1999 Izmit earthquake [4], and for the tectonic loading in southern California [40]. We therefore add to the coseismic slip a deep patch between 20 km and an arbitrary large depth (1500 km). Doing so, we model a thin elastic plate overriding a relaxed lower crust [15,41,42]. Significant relaxed stress change $\Delta \text{CFS}_{\text{rel}}$ then occurs at a larger distance from the Fuyun Fault (Fig. 3c). The Gobi–Altay Fault now suffers a positive (loading) stress of +0.15 bar on its western termination and +0.08 bar on its eastern termination. Since only 26 years separate the Fuyun earthquake and the Gobi–Altay earthquake, the stress relaxation caused by the Fuyun event could not have been complete at the time of the 1957 event. As a consequence, the stress change on the Gobi–Altay Fault should have been smaller than 0.15 bar. Also, our fault geometry neglects local fault complexities. Nevertheless, postseismic stress relaxation seems to have a significant effect between distant faults such as Bolnay, Fuyun and Gobi–Altay faults.

4. Earthquake cycle on the Bolnay, Fuyun and Gobi–Altay faults

We explore the possible consequence of postseismic stress relaxation on the seismic cycle. In modeling this cycle, one needs to generate hundreds or thousands of events in order to statistically quantify the effect of stress transfer on the earthquake timing. Until now, numerical modeling of earthquake series on a fault system has been done to study the generic fault properties, such as for example earthquake predictability [43], faults interaction [44,45] or earthquake activ-
ity mode switching [46]. We attempt here to explicitly model coseismic and postseismic stress interactions between large Mongolian faults and compute earthquake occurrence along a time span.

We consider a model in which each fault is represented by a 1-degree of freedom spring-slider model [47]. Because the 1905 Tsetserleg earthquake probably triggered the Bolnay earthquake 2 weeks later, and because of the rupture complexity, we model the 1905 pair as a unique event. We use the geometry of the largest rupture (Bolnay) in order to compute the interaction with the Fuyun and Bogd faults. Our deterministic model may therefore be viewed as a three-fault system and predict the time of earthquake occurrence on each fault. Using a series of analytic methods, we compute the stress evolution of this three-fault system and predict the time of earthquake occurrence on each fault. Using a series of

\[
\Delta \text{CFS}_i^{\text{CO}}(t) = \begin{cases} 
-0.040 & -0.080 \\
0.001 & -0.000 \\
0.019 & 0.007
\end{cases}
\]  

Bolnay trigger

\[
\Delta \text{CFS}_i^{\text{REL}}(t) = \begin{cases} 
-0.85 & 0.26 \\
0.07 & -0.15 \\
0.73 & 0.22
\end{cases}
\]  

Bolnay trigger

where the line number represents the fault \( j \) where the stress is coming from, and the column number represents the fault \( i \) where the stress is acting on. For example, the value 0.15 bar corresponds to the relaxed stress coming from Fuyun (\( j = 2 \)) and acting on Gobi–Altay (\( i = 3 \)) as shown in Fig. 3c.

We make the simplified assumption that the postseismic stress change on a fault \( i \) coming from an earthquake on a distant fault \( j \) at \( t_f \) evolves from \( \Delta \text{CFS}_i^{\text{CO}} \) at \( t_f \) to \( \Delta \text{CFS}_i^{\text{REL}}(t_f) \) at the end of the relaxation process. If the postseismic relaxation is controlled by a relaxation time \( R \), the postseismic stress for \( t > t_f \) is given by:

\[
\Delta \text{CFS}_i^{\text{POST}}(t) = (1 - e^{-(t-t_f)/R})(\Delta \text{CFS}_i^{\text{REL}} - \Delta \text{CFS}_i^{\text{CO}})
\]  

In order to compute the total interseismic stress \( \tau_i(t) \) acting on a fault \( i \) after an earthquake at \( t_i \), we assume that this stress corresponds to the sum of a secular stress accumulation \( \sigma \), a local stress relaxation after the earthquake, and the sum of stress changes coming from \( N-1 \) other faults. \( \tau_i \) is therefore given by:

\[
\tau_i(t-t_i) = \tau_i^{\text{yield}} + \dot{\sigma} \cdot (t-t_i) - \Delta \tau_i^{\text{co}} \cdot e^{-(t-t_i)/R} - \Delta \tau_i^{\text{rel}} \cdot (1 - e^{-(t-t_i)/R}) + \sum_{j=1, j\neq i}^N \Delta \text{CFS}_j^{\text{CO}} + \Delta \text{CFS}_j^{\text{POST}}(t)
\]  

The total stress \( \tau_i \) increases with time until the yield stress \( \tau_{\text{yield}} \). Earthquake motion then occurs on fault \( i \). The coseismic stress drop \( \Delta \tau_i^{\text{co}} \) is set to 50 bar which is compatible with seismic estimates [25]. Since the stress interaction is already known from \( \Delta \text{CFS}_i^{\text{CO}} \) and \( \Delta \text{CFS}_i^{\text{POST}} \), the free parameters in Eq. 7 remain \( \Delta \tau_i^{\text{rel}} \) and \( R \). We choose a relaxation time of 30 years as proposed by Thatcher [33] for the San Andreas Fault. According to Eq. 3 we choose three values of 5, 10, and 20 bar to encompass possible values of \( \Delta \tau_i^{\text{rel}} \). We suppose a mean recurrence time \( T \) of 4000 years for individual faults [12,13]. The loading rate \( \sigma \) is given by the ratio \( \Delta \tau_i^{\text{rel}}/T \), which leads to secular loading rates ranging from 0.00125 to 0.005 bar/yr. A random variation is introduced in the system with varying \( \tau_{\text{base}} \) in order to encompass the probable scatter of event magnitude on a fault along the seismic cycle. The scalar seismic moment for each fault therefore varies by \( \pm 10\% \) with respect to the value associated to the 1905, 1931 and 1957 events. Integrating Eq. 7 using a finite difference method, we compute the stress evolution of this three-fault system and predict the time of earthquake occurrence on each fault. Using a series of
10,000 earthquakes, we illustrate (Fig. 4) the re-partition of elapsed times $T_{\text{elapsed}}$ between successive earthquakes on this three-fault system. Elapsed times between two successive earthquakes range from 0 to 6000 years, but only short values are displayed. We first tested whether any abnormal recurrence times occur when only the coseismic stress $\Delta \text{CFS}^{\text{CO}}$ (Eq. 4) is used (Fig. 4a). Adding the postseismic stress change (Eqs. 5 and 6 leads to a dramatic increase of short recurrence time. The disturbance is especially marked for small $\Delta \tau_{\text{rel}}$ values (Fig. 4b). The decay of the histogram towards the plateau value occurs for recurrence times between 70 and 100 years, three times as large as the relaxation time $R$.

5. Discussion

Given the geometrical and mechanical simplifications made in our modeling, we do not claim that this analysis may be used in a predictive way. Also, Quaternary seismic activity in Mongolia probably involve tens of faults (Fig. 1), suggesting that various faults can be involved in a clustering process. The interest of the model is rather to reveal that the Bolnay, Fuyun, and Gobi–Altay faults, which are practically uncoupled when considering coseismic motion only, can suffer significant interaction due to postseismic relaxation. As a consequence, a large earthquake may significantly affect the earthquake probability on distant faults for several decades. This time-dependent stress change process is two-fold. First, it reduces the local stress drop where the earthquake took place, bringing the stress level closer to yield limit in only a few decades. This local relaxation favors a possible instability due to a distant trigger. Second, it magnifies the CFS change on distant faults.

The difference between coseismic stress transfer and relaxed stress transfer can be better illustrated using the concept of clock advance (Eq. 2) and the example of the stress transfer from Fuyun to Gobi–Altay (Figs. 2c and 3c). If we consider only coseismic stress change, $\Delta \tau^0$ is in the order of 50 bar [25,26], and $\Delta \text{CFS}^{\text{CO}}$ coming from the Fuyun Fault to the Gobi–Altay Fault is in the order of

![Fig. 4. Histogram of elapsed time $T_{\text{elapsed}}$ between successive earthquakes based on 10,000 earthquake series. Only short values smaller than 200 years are shown. (a) Coseismic stress transfer leads to a flat histogram. This means that the low interfault stress transfer does not promote earthquake triggering between the Bolnay, Fuyun and Gobi–Altay faults. (b) Postseismic stress transfer increases by 10 the abundance of short elapsed times. This time clustering effect is controlled by the relaxation time $R$. A normal abundance is recovered for elapsed times larger than 100 years (c). (d) Larger values of relaxed stress drop lead to a progressive decrease of the clustering effect.](image-url)
0.001 bar. Clock advance is therefore approximately 1 month when considering a recurrence time of 4000 years. Clock advance computations cannot be done directly if postseismic relaxation is being taken into account, since interseismic stress accumulation is no longer linear with time (Fig. 3a,b). However, if we consider periods that display a complete postseismic relaxation, stress accumulation recovery is linear. If the relaxation time is far smaller than the mean recurrence time for individual faults, such as in Mongolia, the relaxed phase lasts most of the interseismic period. During this phase, the loading stress rate is given by scaling \( \Delta \tau_{\text{rel}} \) by \( T \), and the CFS stress change is given by \( \Delta \text{CFS} \). Using \( \Delta \tau_{\text{rel}} = 5 \) bar and \( \Delta \text{CFS} = 0.15 \) bar, we find a clock advance \( \Delta t \) of 120 years. This clock advance is now quite significant, and corresponds to a marked clustering as demonstrated by our seismic cycle model.

Our modeling predicts clusters having a maximum duration of 100 years, i.e. about three times the postseismic relaxation time \( R \). Because this factor of three may also be controlled by the fault number of the tectonic system, we have performed other experiments and have computed this factor using fault systems having lower (two) and higher (six) numbers of faults. Within this fault number range, the cluster duration remains unchanged, which indicates that this duration is mainly controlled by the postseismic relaxation time. According to Eq. 6, such a duration corresponds to an almost complete (95%) postseismic stress effect. In nature, the assessment of this relaxation time relies on a few long-term geodetic measurements after large earthquakes. Because proposed values range between 5 and 30 years [33,48], temporal cluster sizes should be between 15 and 90 years. Due to the low precision of historical and paleoseismic observations, it may be hard to extract reliable earthquake clusters from data sets. Also, local rheological variations may cause scattering of the relaxation time in the lithosphere. However, specific examples indicate cluster durations of approximately 50 years: 52 years for the 1905–1957 Mongolian sequence, 60 years for the 1939–1999 North Anatolian Fault sequence, and \( \sim 50 \) years for the earthquake storm of Nur and Cline [5]. A new cluster may also be the pair 1992 Landers–1999 Hector Mine, which occurred in a slowly deforming area, as attested by the long recurrence time (5–15 years) on the Landers Fault [49]. Therefore, revealing postseismic strain by satellite geodesy [50,31] and modeling earthquake clusters in actively deforming areas could improve our understanding of earthquake occurrence, above all in zones where large earthquakes have already struck in the past decades.

Acknowledgements

This work has been supported by the research program ‘ACI Prévention des catastrophes naturelles’ of the French Ministry of Research. We thank Bettine van Vuuren who greatly enhanced the quality of the English text. We also thank Hugo Perfettini and Ruth Harris for their careful reviews leading to an improved paper.[AC]

References


