Final Technical Report

USGS/NEHRP Award No: G11AP20037 / G11AP20038

Title: Near-Real-Time Monitoring, Analysis, and Forecast of Repeating Earthquakes in Northern California: Collaborative Research with Columbia University and University of Southern California

PIs: Felix Waldhauser[1)], Jeremy Zechar[2)] **Co-PI:** David Schaff[1)]

1) Lamont-Doherty Earth Observatory
Columbia University
Palisades, NY 10964
Tel: (845) 365 8538; Fax: (845) 365 8150
felixw@ldeo.columbia.edu

2) ETH Zurich
Zurich, Switzerland
Tel: +41 44 633 75 74
jeremy.zechar@sed.ethz.ch

Project Period: 01/01/2011 – 12/31/2011 (NFE until 12/31/2012)

Research supported by the U.S. Geological Survey (USGS), Department of the Interior, under USGS award number 08HQGR0043. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Government.

Final Technical Report

USGS/NEHRP Award No: G11AP20037 / G11AP20038

Title: Near-Real-Time Monitoring, Analysis, and Forecast of Repeating Earthquakes in Northern California: Collaborative Research with Columbia University and University of Southern California

PIs: Felix Waldhauser[1)], Jeremy Zechar[2)]   **Co-PI:** David Schaff[1)]

1) Lamont-Doherty Earth Observatory
Columbia University
Palisades, NY 10964
Tel: (845) 365 8538; Fax: (845) 365 8150
felixw@ldeo.columbia.edu

2) ETH Zurich
Zurich, Switzerland
Tel: +41 44 633 75 74
jeremy.zechar@sed.ethz.ch

ABSTRACT Within the scope of this project we have developed and applied tools to systematically search for and identify repeating earthquakes on a large scale across northern California. These events rupture the same fault patch repeatedly, generating virtually identical seismograms. Repeating earthquakes play an important role in the study of fault processes, and have the potential to improve hazard assessment and earthquake forecast. In California, repeating earthquakes have been found predominately along creeping sections of the central San Andreas Fault, where they are believed to represent failing asperities on an otherwise creeping fault.

We used existing databases of 450,000 double-difference locations and over 2 billion waveform cross-correlation measurements as a starting point for the search, and performed additional analysis for clusters of potentially repeating events. These include improved relative location and errors, correlation coefficients over long windows, and differential magnitudes from relative amplitudes. Because of the limited timescale and support of this project we focused on capturing well-defined sequences of 3 or more repeating events between 1984 and 2009. We continue the computationally expensive processing and analysis beyond the scope of this project to expand and update the initial results we generated under this grant.

The preliminary catalog of repeating earthquakes includes 1,876 sequences with 8,612 events in total. The majority of these sequences locate along known faults, while most sequences of 5 or more events locate on faults for which GPS and geologic observations indicate that they are creeping at the surface. The number of events in each sequence ranges from 3 to 24, with the majority of sequences (66%) including less than five events. Mean recurrence times range from less than 1 year (18% of all sequences) to 12 years. Coefficients of variation in recurrence times, CV, range from ~0 to 3, indicating a range of behavior between periodic occurrence ($CV \sim 0$), random occurrence, and temporal clustering.

For the 55 most periodic sequences we conducted retrospective earthquake forecast experiments and calculated hazard functions from fitting four different models to the recurrence intervals. We measured the predictive power of the hazard function using Molchan diagram analysis to conclude that for the 55 sequences the hazard function is significantly more reliable in predicting the next event than random guessing.

1. Overview of Investigations

This final technical report covers the activities performed between January 1, 2011 (start date of the project) and December 31, 2011 (NFE until December 31, 2012). The work described in this report is being undertaken by the principle investigators Felix Waldhauser (Columbia U) and Jeremy Zechar (UCSD/ETH) and co-PI David Schaff (Columbia U).

Within the scope of this project we completed the following two tasks as outlined in the revised version of the originally proposed work plan: 1) Development of a catalog of repeating earthquakes for northern California; 2) Initial analysis of the properties of the sequences of repeating earthquakes. The reduced budget did not allow for developing the tools to dynamically update the repeating earthquake catalog in near-real time.

Repeating earthquakes — earthquakes that rupture the same fault area with sources of similar magnitude and mechanism — are playing an increasingly important role in the study of fault processes and behavior, and have the potential to improve hazard assessment, earthquake forecast, and seismic monitoring capabilities. In Northern California, repeating events of small magnitudes have been found predominantly in creeping zones of the San Andreas Fault system (Poupinet et al., 1984; Vidale et al., 1994; Nadeau et al, 1995; Rubin et al., 1999; Schaff et al., 2002, Waldhauser et al., 2004; Templeton et al., 2008; among others). A popular hypothesis for repeating events therefore is that they represent seismogenic failure of a locked patch on an otherwise creeping fault. However, repeating events have also been found on faults where creep behavior at depth is less well understood, such as the Hayward Fault (Waldhauser and Ellsworth, 2002).

The overall goal of this project was to carry out a comprehensive search for repeating earthquakes across northern California and develop a catalog of repeating earthquakes that encompasses various tectonic settings. The new catalog will form the basis for further analysis and testing of hypothesis on the controling nature of repeating earthquakes. It will also provide the basis for continuous monitoring of the properties of repeating earthquakes, such changes in the recurrence intervals, from which changes in the loading rate may be inferred.

Due to the limited timescale of this project we restricted our search to sequences with 3 or more repeating events, and for the time period 1984-2009. We continue to expand this preliminary catalog beyond this project's end date to include sequences with two events and analyze sequences that were likely missed during the restrictive first stage screening procedure described below. We also intend to update the catalog with more recent years, and eventually detect and associate new repeats as part of the DDRT monitoring effort (Waldhauser, 2009). In the following we show results from this first step towards a comprehensive dynamically updated repeating earthquake catalog for Northern California.

2. Investigations undertaken

2.1 Data and search procedure

We have developed a search procedure for identifying sequences of repeating earthquakes in northern California. As starting point we used the northern California double-difference catalog of 450,000 precisely located events (1984-2009) (Waldhauser and Schaff, 2008) and associated database of over two billion waveform cross-correlation coefficients (Schaff and Waldhauser, 2005). The correlation database includes P- and S-wave cross-correlation coefficients ≥ 0.7 for all pairs of events separated by less than 5 km and observed at common stations. The time-domain correlation measurements were carried out within a 1 s window around the P-wave arrival (2 s windows for S-waves) on filtered (1.5-15 Hz) short period seismograms from the NCSN archive (for details see Schaff and Waldhauser, 2005). Approximately 90% of the events in the double-difference catalog have correlated P- or S-wave trains with at least one neighboring event (Waldhauser and Schaff, 2008).

We searched the correlation database for event pairs that have at least one P-wave correlation coefficient ≥ 0.95 and a mean of the five highest P-wave coefficients ≥ 0.95. From the 700,025 resulting event pairs with a total of 164,592 unique events we selected the pairs that have events separated by less than the approximate source radius (calculated from a 3 MPa constant stress drop source) and a magnitude difference less then 0.3 units (based on the NCSN magnitudes). We then grouped pairs with common events into 12,082 clusters, including a total of 41,418 events. At this point we only considered clusters with three or more events for further processing and analysis (i.e., 4,611 clusters with a total of 26,574 events). The spatial distribution of events in these clusters range from virtually identical locations (Figure 1a) to tightly clustered events that appear to occupy overlapping fault areas (Figure 1b). In order to add more constrain on the characteristics of these sequences and separate the former from the latter we performed additional analysis.

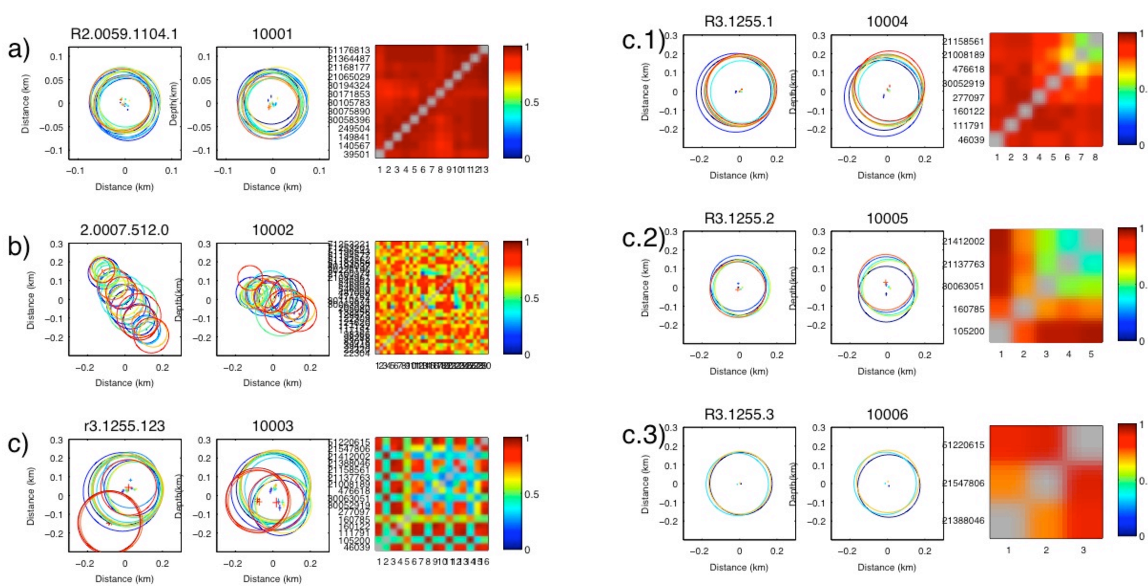

Figure 1 *Examples of clusters of tightly located and correlated earthquakes found by the search procedure. A) co-located; b) spatially separated; c) three individual sequences that appear to occupy part of each other's rupture area. c.1-3) Each of the 3 sequences in c) individually analyzed.*

We relocated each cluster individually to resolve the fine details in the hypocentral locations and to compute formal least-squares relative location errors using the double-difference algorithm *HypoDD* (Waldhauser and Ellsworth, 2000; Waldhauser, 2001). The horizontal and vertical least-squares location uncertainties at the 90% confidence level are on average 4 and 8 m, respectively. For comparison, the corresponding NCSN locations deviate on average 97 m horizontally and 225 m vertically from the cluster centroid.

We further computed cross-correlation coefficients within windows that capture most of the wave train, starting 1 sec before the P- and ending 5 sec after the S-wave arrivals. Compared to the short window correlations, which are aimed at optimizing the alignment of the phase onsets for delay time measurements and precise location, the long-window correlations are more useful for comparing overall seismogram similarity between events in a given cluster.

Finally, we computed relative magnitudes by measuring cross-correlation based relative amplitudes between pairs of seismograms at common stations (Schaff and Richards, 2011). We assume for the repeating events that the slave event is a scaled, identical version of the master event in the presence of noise. The least-squares solution for the amplitude factor is equivalent to the unnormalized cross correlation coefficient divided by the inner product of the master waveform. Differences between relative magnitudes derived from coda-duration techniques employed by the NCSN and our correlation-based estimates follow a Gaussian distribution (with longer tails) and have a standard deviation of 0.09 magnitude units (Figure 2). The standard deviation is greater for smaller events than it is for larger ones. In all subsequent analysis we use the correlation based magnitude estimates.

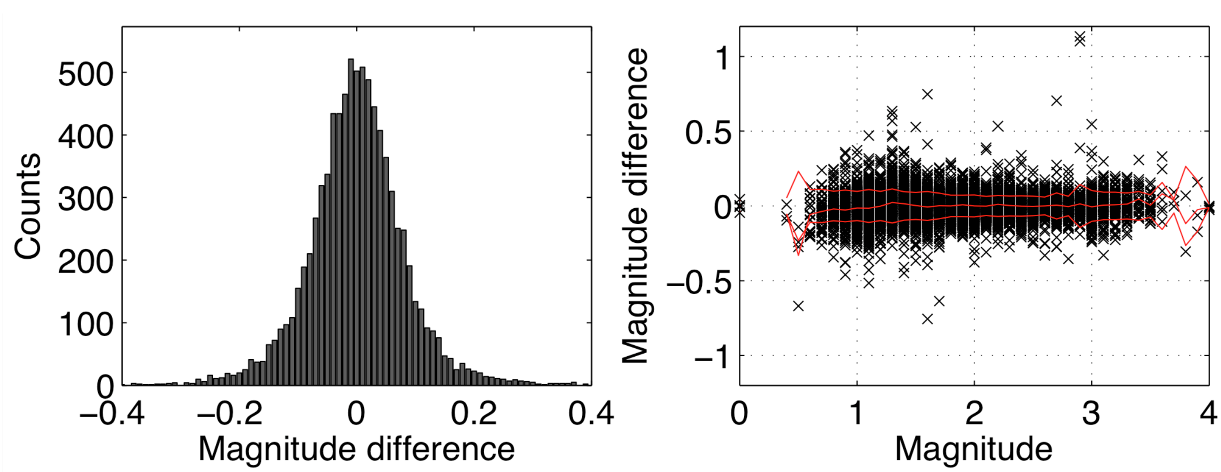

Figure 2 *a) Histograms of difference between NCSN and correlation based magnitudes. b) Distribution of differences as a function of magnitudes. Red lines indicate standard deviation.*

The additional measurements were used to identify sequences of repeating events from the 4,611 clusters of strongly correlated and tightly located events. We followed Waldhauser and Ellsworth (2004) and use the following criteria for identification:

1) All events in a sequence have to be within each other's 3 MPa stress drop rupture area, accounting for the formal error in the relative magnitude estimation and the relative location errors of the hypocenters.

2) Only events that are within 0.3 magnitude units of the median magnitude of all events in a given sequence are considered.

3) Each event has to be linked to at least one other event in the sequence via high correlation coefficients as determined from wave-train correlation.

We manually inspected each sequence that passed the search and fixed problems not captured by the search procedure. One such problem is shown in Figure 1c where 3 apparently distinct sequences nucleate from within each other's rupture area. We separate these to build 3 sequences (Figure 3c.1-3). We note here that it is also possible that these 3 sequences rupture the same fault surface but nucleate from different locations within that surface. The assumption of a 3 MPa circular rupture in our search procedure may not be appropriate in such cases.

2.2 Results

The preliminary catalog of repeating earthquakes includes 1,876 sequences with a total of 8,612 events (Figure 3). The majority of these sequences locate along known faults, with sequences of 5 and more events locating on faults for which GPS and geologic observations indicate that they are creeping at the surface. The number of events in each sequence ranges from 3 to 24, with the majority of sequences (66%) including less than 5 events and 102 sequences (5.4%) having 10 or more events (Figure 4a). The average recurrence time, calculated from the times between subsequent events in each sequence, range from less than 1 year indicating burst like behavior (18% of the sequences) to 12 years (Figure 4b). Coefficients of variation, *CV*, calculated for each sequence by dividing the standard deviation of the recurrence times, *Tr*, by their mean,

$$CV = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Tr_i - \overline{Tr}\right)^2}}{\overline{Tr}}$$

are between ~0 and 3 (Figure 3c), indicating a range of behavior between periodic occurrence ($CV \sim 0$), random occurrence, and temporal clustering.

The temporal behavior of the repeating earthquakes can be broadly characterized by four types: (1) periodic occurrence (CV ~ 0) (Figure 5a); (2) random occurrence (CV>>0) (Figure 5b); (3) temporal clustering (Figure 5c); and (4) piecewise periodic occurrence (Figure 5d). We have focused some of our efforts on analyzing the predictive power of the most periodic sequences.

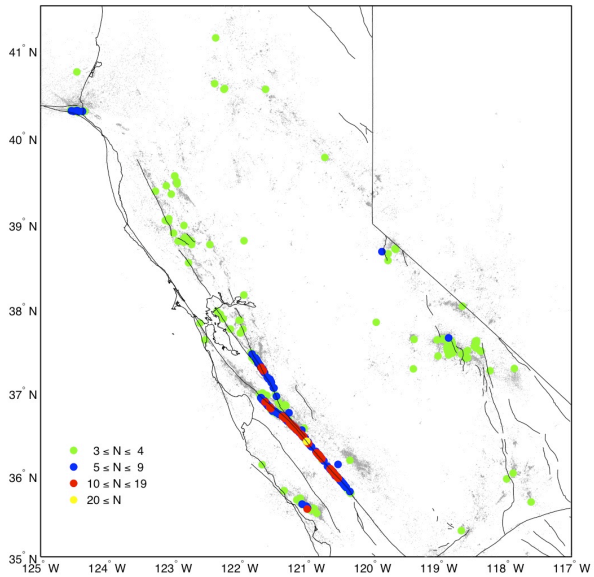

Figure 3 *Distribution of sequences with 3 or more events resulting from the initial search. We continue to update this catalog by analyze additional sequences, including the ones with only two events, that show all characteristics of repeating earthquakes but did not pass through our first stage screening process.*

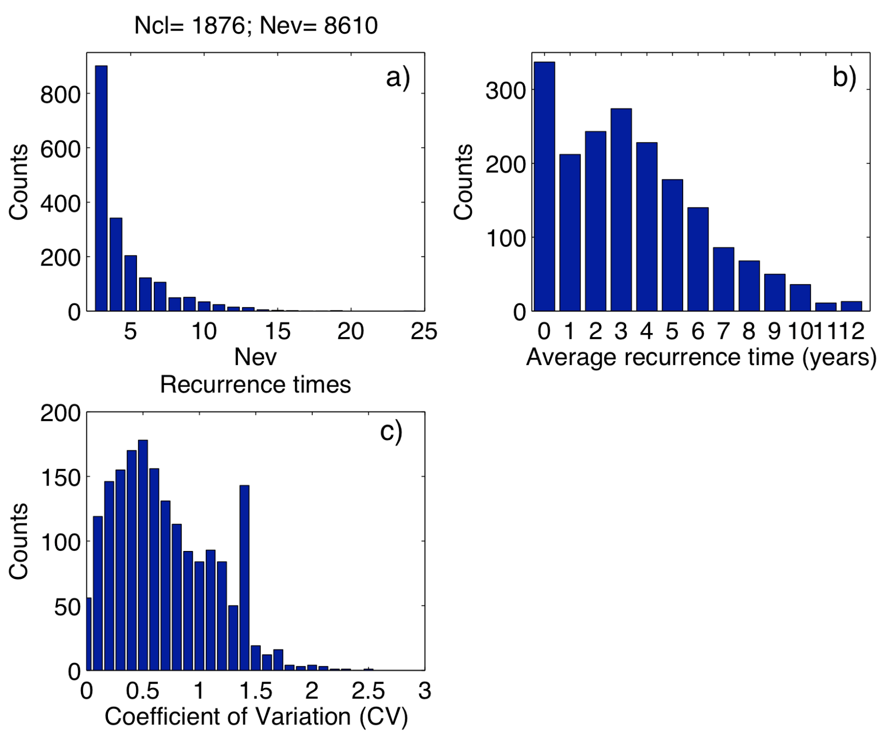

Figure 4 *Histograms showing the distributions of a) the number of events, b) average recurrence times, and c) the coefficients of variation of the recurrence times in each sequence.*

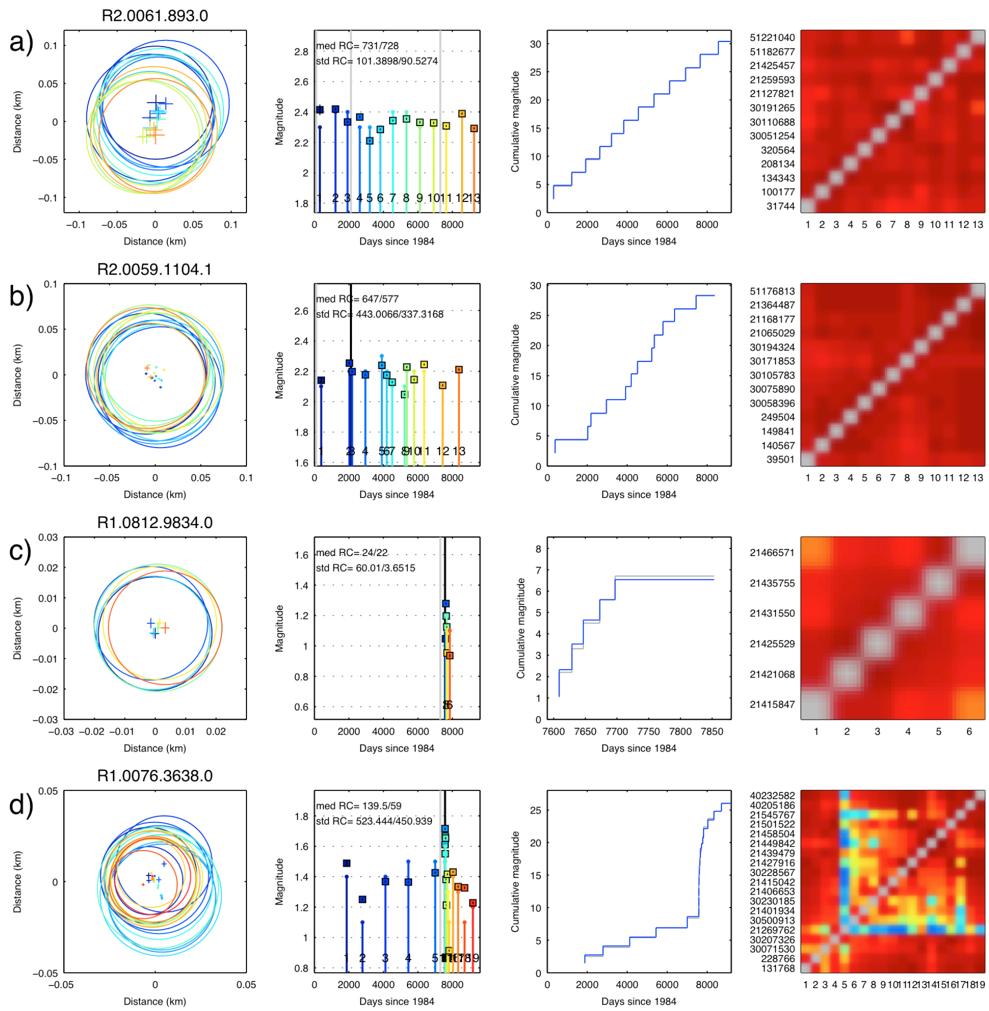

Figure 5 *Spatio-temporal characteristics of 4 selected repeating earthquake sequences (a-d). Shown are location of events represented by circles scaled by a 3 MPa stress drop source (1st column of panels), time and magnitude of events (2nd columns), cumulative magnitudes as a function of time (3rd panels), and matrix of correlation coefficients.*

2.3 Predictive power of periodic sequences

We explored the predictability of 55 "clean" sequences with at least 3 repeats; these sequences met the following criteria:

- Recurrence time median greater than 30 days;
- Recurrence time sample standard deviation less than 365 days; and
- Relative magnitude sample standard deviation less than 0.06.

Figure 6 can be compared with Figure 4 to show how these sequences are different from the population of sequences.

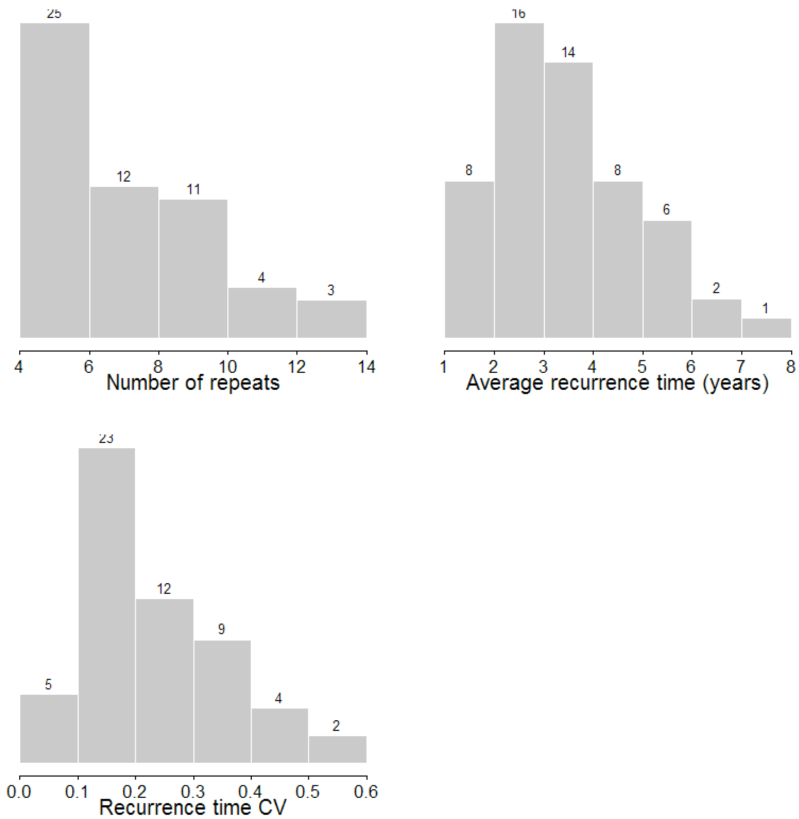

Figure 6 *Same as Figure 4 but only considering the 55 clean sequences of 4 or more events.*

Although we didn't directly use recurrence time coefficient of variation to identify clean sequences, using a maximum sample standard deviation and minimum median has a similar effect: the CV distribution for these sequences is very different from the general population.

With these sequences, we conducted retrospective earthquake forecast experiments. For each sequence, we began by considering the first three events in the sequence. We fit a Weibull, lognormal, exponential, and inverse Gaussian model to the corresponding two recurrence intervals and determined which fit best according to Akaike Information Criterion (Akaike, 1974). With the best model among these candidates chosen, we calculated the corresponding hazard function for each 24-hr interval after the second event until the third event. After each subsequent event, we checked to see which of the four models now fit the data best and updated the hazard function.

The hazard function is an example of what Zechar & Jordan (2008) called an alarm function, any ordering function that indicates the *relative* propensity for a future earthquake. In other words, when an alarm function increases, this implies a predictive statement that earthquake propensity is increasing. An advantage of an alarm function analysis is that it can be applied to nearly any candidate precursory signal and does not require a probabilistic statement. We measured the predictive power of the hazard function using Molchan diagram analysis. With this method, we consider the 24-hr hazard function values as a time series, and we assess whether the hazard function values are high just before the repeats happen. To do this, we choose a threshold value and consider all hazard function values above this threshold to be alarms—that is, a statement that a repeat will happen tomorrow. All values below the threshold are not alarms. We consider two metrics: alarm time fraction and miss fraction. The alarm time fraction is the number of alarms divided by the number of days in the experiment; in this case, the experiment lasts from the date of the third repeat until 2010 (the end of the catalog). The miss fraction is the fraction of repeats that occur on days when no alarm is declared. In the perfect situation, an alarm function would obtain a miss fraction of zero and a trivially small alarm time fraction. For a specific threshold value, one can plot the paired couple of alarm time fraction versus miss fraction. Repeating this process for all possible threshold values (in this case, from the maximum hazard function value to zero) traces out a monotonically decreasing Molchan trajectory. If the alarm function has no reliability—i.e., if someone is declaring alarms at random, the Molchan trajectory will, on average, follow the diagonal from (0, 1) to (1, 0). In Figure 7, we show the Molchan trajectory for our method, upon stacking all sequences.

Beyond the visual indication that this method is better than random guessing (i.e., the majority of squares fall outside the 95% confidence bounds for the reference of random guessing), we have a quantitative measure—the area skill score. The area skill score, described by Zechar & Jordan (2008, 2010), is the area above the Molchan trajectory and for random guessing its expected value is 0.5. We performed a statistical hypothesis test under the assumption that the area skill score A follows a normal distribution with a variance that depends on the number of earthquakes, n. In particular, we use the results of Zechar & Jordan (2010): $A \sim N(\mu, 1/12n)$. The null hypothesis is that $\mu = \mu_0$ where μ_0=0.5, and the alternative hypothesis is $\mu > \mu_0$. With a significance value of 0.01 and n=230, the null hypothesis rejection region is any value greater than 0.5442. For the 55 sequences considered here, the area skill score is 0.7213, corresponding to a p-value that is numerically indistinguishable from zero. Therefore, we reject the null hypothesis, and in doing so conclude that the hazard function is significantly more reliable than random guessing.

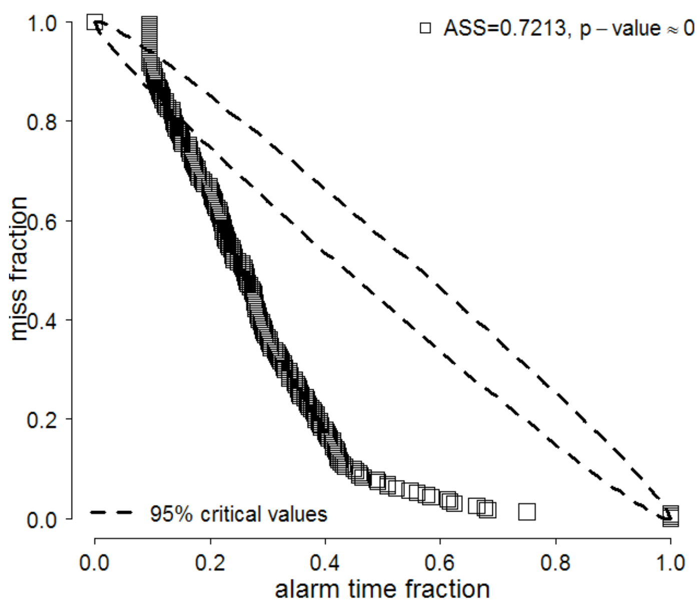

***Figure** 7 Molchan trajectory indicating predictability of the 55 clean sequences. Each square corresponds to a unique threshold value.*

These results are of course influenced by the selection bias we introduced by filtering the sequences, but we include this example only as a proof-of-concept. An application to all sequences, extension to probabilistic forecast models and evaluation, and exploration of these sequences in the context of operational earthquake forecasting could be future avenues of research, particularly for researchers participating in the Collaboratory for the Study of Earthquake Predictability.

3. References

Akaike, H., A new look at the statistical model identification, *IEEE Trans. Autom. Control*, 19, 719–723, 1974

Nadeau, R.M., W. Foxall, and T.V. McEvilly, Clustering and periodic recurrence of microearthquakes on the San Andreas fault at Parkfield, *Science*, 267, 503-507, 1995.

Poupinet, G.,W.L. Ellsworth, and J. Fréchet, Monitoring velocity variations in the crust using earthquake doublets: an application to the Calaveras fault, California, *J. Geophys. Res.*, *89*, 5719-5731, 1984.

Rubin, A.M., D. Gillard, and J.-L. Got, Streaks of microearthquakes along creeping faults, *Nature, 400*, 635-641, 1999.

Schaff, D.P., G.H.R. Bokelmann, G.C. Beroza, F. Waldhauser, and W.L. Ellsworth, High resolution image of Calaveras Fault, California, seismicity, *J. Geophys. Res.*, 107, 2186, doi:10.1029/2001JB000633, 2002.

Schaff, D.P. and P.G. Richards, On finding and using repeating seismic events in China, *J. Geophys. Res.*, 116, B03309, doi:10.1029/2010JB007985, 2011.

Schaff, D.P. and F. Waldhauser, Waveform cross-correlation-based differential travel-time measurements at the Northern California Seismic Network, *Bull. Seism. Soc. Am.*, *95*, 2446-2461, 2005.

Templeton, D.C., R.M. Nadeau, and R. Buergmann, Behavior of repeating earthquake sequences in Central California and the implications for subsurface fault creep, *Bull. Seism. Soc. Am., 98,* 52-65, 2008.

Vidale, J.E., W.L. Ellsworth, A. Cole, C. Marone, Variations in rupture process with recurrence interval in a repeated small earthquake, *Nature, 368,* 624-626, 1994.

Waldhauser, F., and W.L. Ellsworth, A double-difference earthquake location algorithm: method and application to the northern Hayward Fault, California, *Bull. Seism. Soc. Am., 90,* 1,353-1,368, 2000.

Waldhauser, F. and W.L. Ellsworth, Fault structure and mechanics of the Hayward Fault, California, from double-difference earthquake locations, *J. Geophys. Res.*, 107(B3), 2054, doi:10.1029/2000JB000084, 2002.

Waldhauser, F., W.L. Ellsworth, D.P. Schaff, and A. Cole, Streaks, multiplets, and holes: High-resolution spatio-temporal behavior of Parkfield seismicity, *Geophys. Res. Lett. 31*, L18608, doi:10.1029/2004GL020649, 2004.

Waldhauser, F. and D.P. Schaff, Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, *J. Geophys. Res.*, 113, B08311, doi:10.1029/2007JB005479, 2008.

Waldhauser, F., Near-real-time double-difference event location using long-term seismic archives, with application to Northern California, *Bull. Seism. Soc. Am.,* 99, 2736-2848, doi:10.1785/0120080294, 2009.

Zechar, J.D. and T.H. Jordan, Testing alarm-based earthquake predictions, *Geophys. J. Int.*,

172(2), 715-724, doi:10.1111/j.1365-246X.2007.03676.x, 2008
Zechar, J.D. and T.H. Jordan, The area skill score statistic for evaluating earthquake predictability experiments, *Pure Appl. Geophys.*, 167 (8/9), 893-906. doi:10.1007/s00024-010-0086-0, 2010.

4. Reports published related to this project, or previous related projects

Peer-reviewed journal publications:
In preparation.

Conference/meeting abstracts:

Waldhauser, F., D.P. Schaff, J.D. Zechar, B.E. Shaw, Distribution and Characteristics of Repeating Earthquakes in Northern California, Abstract S31A-2487, Fall Meeting, AGU, San Francisco, Calif., 3-7 Dec, 2012.

Waldhauser, F., J.D. Zechar, D.P. Schaff, and P. Friberg, Real-Time Double-Difference Location and Monitoring of Fine-Scale Seismogenic Properties, with Application to Northern California, EGU General Assembly, Vienna, 2012-11630, 2012.

Waldhauser, F., D.P. Schaff, J.D. Zechar, P. Friberg, Real-time double-difference location and monitoring of repeating earthquakes in Northern California, SSA Annual Meeting, *Seism. Res. Lett.* 83, 2, 2012.

*Waldhauser, F., D. Schaff, B. Engebreth, P. Friberg, and J.D. Zechar, Real-time double-difference locations at the NCSS and monitoring of repeating earthquakes, 9th Annual Northern California Earthquake Hazards Workshop, Menlo Park, CA, January 24-25, 2012.

*Waldhauser, F., D. Schaff,B. Engebreth, P. Friberg, Towards Real-Time Monitoring of Fine-Scale Seismogenic Properties in Northern California, 8th Annual Northern California Earthquake Hazards Workshop, Menlo Park, CA, January 24-25, 2011. (KEYNOTE)

5. Available Data and Products

For inquiries and requests:

Contact: Dr. Felix Waldhauser
Lamont-Doherty Earth Observatory
61 Route 9W, Palisades, NY 10964
Tel. 845.365.8538
felixw@ldeo.columbia.edu

6. Final Remarks

This grant supported the development of search procedures for the generation of a comprehensive catalog of reating earthquakes in northern California. Because of the limited time frame of the project we focused our efforts on a subset of the data. Specifically, sequences with two events were not considered, and the search was carried out using the seismic archive up to the year 2009. Ongoing and future work is aimed at including all sequences of repeating events, and at developing an automatic ‘update’ procedure that searches the archives of subsquent years and eventially new events in real-time using the DD-RT system for members of existing sequences in the base catalog.