Small Earthquake Detection Through Seismogram Comparisons

This summer I have the pleasure of traveling to Stanford and conducting seismological research with Professor Greg Beroza and USGS researcher Bill Ellsworth. In this project we are working to analyze large earthquake data sets and trying to develop novel algorithms to speed up the computations. I am deeply enthusiastic about this project because I am trained in math and statistics and I feel I can really use my skills to make a contribution. Moreover, I am given a great deal of freedom in this project and I believe this is a great opportunity to exercise my creativity in the scientific arena.

The goal of this project is to ultimately look for repeating earthquakes and to uncover low frequency earthquakes, specifically those along the San Andreas fault. Repeating earthquakes are ones that produce very similar seismograms. The figure below shows 10 earthquakes recorded by a Parkfield station. Their similarity indicates their proximity to one another and the type of earthquakes we may expect along this fault. Low-frequency earthquakes are typically magnitude 3 or below and occur in depths of 25 km or greater (whereas larger quakes are occur in depths between 10 and 12 km). These smaller quakes often go undetected in seismological catalogs and hunting down these elusive events may shed light into larger, deadlier earthquakes.

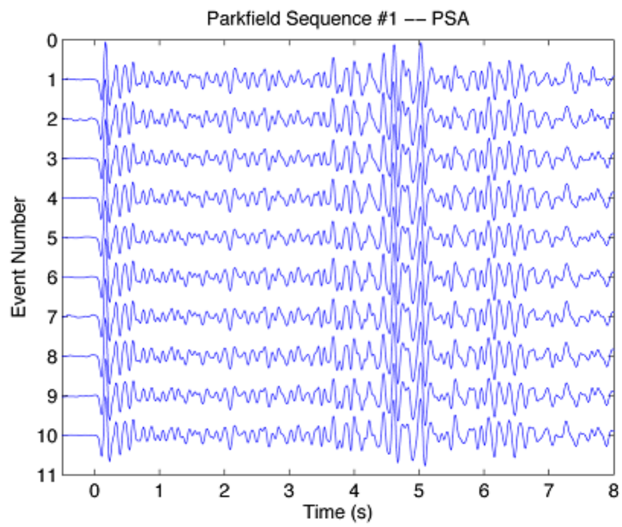

Figure courtesy of USGS and Bill Ellsworth.

The techniques we are using come from statistics and signal processing. We first extract data from the Northern California Earthquake Data Center website. Next we use coefficient correlations (which gauges how closely two sets of data sets are) as a benchmark for comparison. However, we are planning to use more sophisticated statistical techniques such as the Kolmogoroff -- Smirnoff test, which can check whether the correlation coefficients fit our hypothesized probability distribution.

One issue in dealing with seismological data is the abundance of noise. The data recorded may not reflect earthquake waves but rather extraneous events such as wind, vehicle traffic, and weather conditions that may have triggered the sensitive equipments. However, the matched filter technique we use should be able to optimize the signal-to-noise ratio, which can reduce unwanted signals and help us detect real quake tremors. So far I have been doing point-to-point comparisons which I use a small segment of one seismogram and "scanned" it across another and compute the correlation coefficients between the two sets of data segment being tested. Since we will be working with very large data sets such comparisons will be computationally intensive, we want to move toward a more efficient comparison method such as comparing every other data point or randomly selecting points to compare. While my current computations are being programmed in R, a language for statistical computing, hopefully they can be ported into low-level languages such as Fortran or C to speed up the computations even more. Indeed a great deal time is spent developing ways to budget computation time and CPU usage but hopefully our efforts will allow future researchers to scour for repeating earthquakes which may lead to findings on bigger quakes.

- *Raymond Chen*