

Characterizing Seismic Swarm Morphology

Sutton Chiorini

04/25/16

Advisor: Prof. Vedran Lekic

GEOL394

Abstract

Seismic swarms are characterized by an anomalously large number of earthquakes occurring in a relatively small area, typically ranging from a few to several kilometers, over a short period of time, typically ranging from days to weeks. However, how and why swarms occur is poorly understood, which poses an interesting set of questions within the greater body of geologic research. In this study, I propose that previous methods of identifying seismic swarms from larger bodies of earthquake catalogs are not effective in characterizing the full range of possible swarm behaviors. Furthermore, I propose that if a full seismic swarm catalog were to be compiled and analyzed, the sequences will demonstrate a much more even distribution in space and that a larger fraction of swarms will be shown to migrate in both time and space. Finally, I propose that over time, a smaller fraction of swarms will exhibit an exponential decay in event occurrence over time.

I will start my research by using a set of analyses implemented in MATLAB to analyze two different methods for isolating seismic swarms from larger earthquake catalogs. The first method is a threshold analysis proposed by Vidale and Shearer (2006), and the second is a clustering analysis proposed by Zaliapin et al. (2008). I will then demonstrate that although both methods do identify seismic swarms, the former method does not identify all the seismic swarms identified by the latter, and the seismic swarms it does identify contain less events than the seismic swarms identified by the former. I will then discuss why this might be the case by analyzing the spatial distribution, migration characteristics, and decay behavior of all the seismic swarms that the Vidale and Shearer method missed, in order to more accurately characterize the full range of observed swarm behavior. Using this understanding, I will expand Vidale and Shearer's thresholds to create a complete catalog of earthquake swarms in California. Finally, I will conclude by comparing the original and adjusted thresholds using a variety of mathematical methods in order to demonstrate that the adjusted thresholds identify more swarms and more events overall, and therefore lend a more complete description to the morphology of swarms in California.

1) Introduction

Seismic swarms, sequences of earthquakes occurring in a relatively small area (between kilometers to tens of kilometers) over a relatively short period of time (between days to weeks), are defined in a variety of ways in the seismological literature. Corral (2003) defines swarms as seismic activity that is not associated with a main event, and therefore do not follow Omori's law, in which the frequency of aftershocks decreases approximately with the inverse of time after the main shock. Vidale and Shearer (2006; henceforth VS2006) define swarms to include many earthquakes striking in a limited space over a limited time window. Using the VS2006 approach to analyze swarms in Southern California, Chen and Shearer (2011) conclude that seismic swarms are thought to be mainly triggered by physical processes. Zhang and Shearer (2016) use a slightly different approach to classify swarms in the San Jacinto Fault Zone, concluding that swarms specific to that location are better explained by fluid flow.

Although Zaliapin et al. (2008; henceforth Z2008) does not propose a specific definition for seismic swarms, they do detail that clustered events they have identified are much closer to each other in time and space than would be expected for a uniform Poisson distribution. In addition,

Zaliapin and Ben-Zion (2013) further demonstrate that swarms can be more broadly classified as one of two dominant types of small-medium earthquake families, the other being burst-like sequences.

Seismic swarms are generally composed of low-magnitude events, the largest of which generally does not exceed M_w 5.0, and on average events run between M_w 0.0 and 2.0. Because of their small size, swarm earthquakes do not pose an immediate threat to surrounding communities and infrastructure upon initiation of the sequence. Nevertheless, swarms have been shown to be associated with magma or fluid activity (Ruppert et al. 2011), as well as stress loading in fault areas and slow aseismic slip events (Vidale and Shearer, 2006, Lohman and McGuire, 2007). Therefore, swarms are related to the processes that result in volcanic eruptions and stress accumulation that can lead to large, destructive earthquakes. Mechanisms for where and why swarms occur are not well understood, and can vary depending on geologic and tectonic setting. Thatcher and Brune (1971) suggest that swarms can be constrained to normal fault settings along mid ocean ridge spreading in the Gulf of California, while Parotidis et al. (2003) propose that seismic swarms in NW Bohemia are triggered due to pore pressure perturbations. Therefore, an improved understanding of seismic swarms can elucidate not only the relationship between swarms and geologic processes, but also point to yet other triggering mechanisms for seismic swarms that have not been explored yet.

In addition, seismic swarms have been demonstrated to have a variety of applications to other geological questions. Savage et al. (2015) showed that seismic swarms can be used as a predictor of volcanic eruptions, providing an additional method for hazard mitigation in volcanic areas. Umeda et al. (2015) demonstrated that seismic swarms can precede a megathrust event, in their specific case the 2011 Tohoku earthquake. As this event was extremely damaging both in terms of lives and infrastructure, this could offer a new way to predict and avoid another such disaster. Finally, Shapiro and Dinske (2008) demonstrated seismic triggering due to fluid injection exhibits swarm-like behavior, which could help in investigations regarding seismic activity associated with hydraulic fracturing. All of these methods demonstrate that better understanding how and why seismic swarms could provide a fruitful source of information for other geological investigations.

Earthquake catalogs are large and contain hundreds of thousands of events for a seismically active region like Northern California over the course of three decades. Therefore, identifying what pattern is a swarm and what is simply background seismicity is a complicated task with no straightforward methodology. Geophysicists have identified a variety of methods by which to identify seismic swarms from large catalogs of earthquakes. Previous research has identified seismic swarms specific to volcanic regions (Fischer et al. 2003; Ruppert et al. 2011), and presented methods for identifying seismic swarms (Vidale and Shearer 2006; Zaliapin et al. 2008; Chen and Shearer 2011) in non-volcanic settings. I will base my research on the methods proposed by Vidale and Shearer (2006) and Zaliapin et al. (2008). The first research method, VS2006, uses arbitrary-chosen parameters defining spatial, temporal, and earthquake count thresholds to identify 71 seismic bursts in Southern California. The second method, Z2008, uses a statistical methodology to analyze spatio-temporal distributions of earthquakes and identify anomalously clustered sets within the population of all earthquakes.

Vidale and Shearer (2006)

Vidale and Shearer picked seismic bursts using 166,525 events from the waveform relocated catalog SHLK_1.01 in Southern California based on an arbitrary-chosen set of criteria, which consists of the following: 1) An initial event must be followed by at least 39 events within a radius of 2 km in 28 days; 2) There must be 3 or fewer events in the prior 28 days within the same 2 km radius; and, 3) No more than 20% more events can occur between 2 and 4 km from the initiating event in the same 28 days afterwards.

Based on these threshold parameters, Vidale and Shearer identified 71 seismic clusters. Fourteen of these bursts resembled main shock/aftershock sequences, eighteen exhibited “swarm-like” sequences or behaviors lacking a clear main shock, and 39 were considered to be “average” sequences, falling between the mainshock/aftershock and swarm sequence behavior. Vidale and Shearer then analyzed the eighteen swarms identified using their particular choice of thresholding parameters (Figure 1), and concluded that seismic swarms tended to exhibit an interval of steady seismicity rate, and that the largest event in the swarm tended to strike later in the sequence. They also found a weak correlation between the number of events in each burst and the magnitude of the largest event in each burst, and that shallow sequences were most likely to be swarm-like. Finally, they found that swarms were most likely to occur in extensional (normal faulting) settings. They concluded that seismic swarms were most likely driven by pore fluid pressure fluctuations and that they are most likely a general feature of tectonic faulting, rather than specific to volcanic or geothermal regions.

Despite its interesting findings and influence on the swarm community (the study has been cited 99 times according to Google Scholar), Vidale and Shearer’s analysis of seismic swarms is not without limitations. Their use of an arbitrarily-chosen set of parameters to define what is and is not a swarm makes it difficult to quantify how many swarms might have been missed by their analysis. Therefore, whether the conclusions of their analysis can be generalized to all swarms remains an open question. In addition, their sample set of seismic swarms, limited to eighteen bursts, is a fairly small sample size, making it difficult to generalize the characteristics they identify to all swarms globally, or even throughout California, because the tectonics of Southern California differ substantially from those in Northern California. Further research on swarm sequences carried out using an objective set of parameters to define a swarm, and done outside the Southern California setting, would help answer these questions.

Zaliapin et al. (2008)

Zaliapin et al. proposed using a statistical methodology for analyzing the clustering of seismicity in the time-space-energy domain. They established the existence of two statistically distinct populations of earthquakes: clustered and non-clustered. Clustered earthquakes can be considered to belong to a swarm population, and the non-clustered events to a Poisson or non-swarm population. This method was developed based on the analysis of Baiesi and Paczuski (2004), and is built upon the parameter η_{ij} , or pairwise earthquake distance in space, time, and energy; the smallest η_{ij} across all j is associated with a particular earthquake i . The catalog used in their research was produced by the Advanced National Seismic System.

Zaliapin et al. first identified the location, time, and magnitude of each event in their catalog, and then used those parameters to calculate an interoccurrence time (T) between pairs of events, denoted by subscript i and j , as well as a spatial distance (R) between the two events. They then normalized R_{ij} and T_{ij} by the magnitude of earthquake i . By multiplying these two parameters, they produced a nearest neighbor distance η_{ij} . A scatter plot of values of R_{ij} and T_{ij} corresponding to the minimum η_{ij} for each earthquake can then be used to identify two statistically distinct earthquake populations (Figure 2a): A population of earthquakes that follow a Poisson distribution (background seismicity) found at larger distances from the origin and tracing out a field with linear and negatively sloped distribution in the R - T space; A population of clustered seismicity (including swarms) at smaller distances from the origin. Histograms of the values of η (Figure 2b) follow a bimodal distribution between the clustered (smaller average η) and non-clustered (larger average η) populations of earthquakes (Figure 3). Based on this analysis, Zaliapin et al. concluded that seismic swarms could be identified using their clustering in the spatial-temporal-energy domain.

As it is based on statistical analysis of the clustering of seismicity in the energy-normalized spatio-temporal domain, the Zaliapin et al. method dispenses with the need for using arbitrary threshold parameters to identify swarms. Nevertheless, the method does have some limitations. Rather than identifying specific bursts of seismicity as Vidale and Shearer did, Zaliapin et al. identifies all swarm events without identifying the swarm they belong to. In other words, although all the swarm events in a catalog can be identified using this approach, there is no way to tell which event belongs to which swarm, or if events belong to multiple swarms. Therefore, events could be misidentified with an unrelated earthquake swarm if they happen to coincide in time and space with that swarm.

2) Hypothesis and Proposed Work

Despite the fact that a number of methods have been established to identify seismic swarms from earthquake catalogs, there is no clear procedure for determining whether all the seismic swarms in an earthquake catalog have been identified, or if all events being identified as swarm events are unique to a swarm. As a result of these difficulties, to date, there are no comprehensive compilations of swarm earthquakes, even though complete catalogs of earthquakes have existed for decades.

I propose to compile a comprehensive catalog of swarm earthquakes – as opposed to the incomplete swarm compilations that are typically analyzed – and analyze the spatial and temporal distributions of the swarm events. I hypothesize that these distributions will not fully match the conclusions advanced by previous investigators. More specifically, I propose a two-part hypothesis: 1) that the swarms I identify will not exhibit an exponential decay curve in terms of a normalized distribution over time, and that they will exhibit a much more even distribution in space than was found by Vidale and Shearer (2006); and, 2) that a larger fraction of swarms will be shown to migrate than found by Chen and Shearer (2011). I will use methods proposed by Vidale and Shearer (2006) and Zaliapin et al. (2008) to analyze two hypocenter double difference catalogs compiled by Waldhauser (2013) for Northern California and Hauksson, Yang

and Shearer (2011) for Southern California in order to compile and analyze the results, morphology, number of events and locations of my seismic swarm catalog.

I expect that the significance of my work will add to the understanding of seismic swarm mechanisms, regarding both where and why they occur. In addition, my work will build upon previous research done by Vidale and Shearer (2006) and Zaliapin et al. (2008) by expanding their categorization of seismic bursts and swarm events in the form of a complete swarm catalog. Finally, my research will add to the existing body of information regarding the temporal and spatial distribution of swarms as they migrate in space and time, which in turn will help categorize their structure and possibly elucidate their relationship to other geologic processes.

3) Methods

HypoDD Algorithm and Resulting Catalogs

Waldhauser and Schaff (2008) developed a catalog of 513,474 events in Northern California spanning 27 years, from the digital seismic waveform archives of Northern California. The catalog is produced using both waveform cross correlation and double difference methods, in which pairs of events with correlated waveforms are then inverted for the precise relative locations of events using a hypocenter double difference (*hypoDD*) algorithm. This method relates the observed and the predicted travel-time differences for pairs of earthquakes observed at common stations to their hypocenters in order to link them through a chain of nearest neighbors, resulting in a high-resolution relative hypocenter locations over a large area (Figure 4). Hauksson, Yang and Shearer (2011) apply a similar three-dimensional velocity model in order to locate Southern California seismicity from 1984-2011 (Hauksson, Yang and Shearer, 2012), within the same time period as the catalog produced by Waldhauser.

For identification and analysis of seismic swarms, I modified and wrote analysis codes in MATLAB based on the work of Vidale and Shearer, and Zaliapin et al. For the former, I used a program written by Jeff Gay that applies the five parameters identified by VS2006 in order to isolate all events that fit within those specified values. For the latter, I wrote a program based on Z2008 that calculates an interoccurrence time, distance between all events, normalizes both parameters based on the magnitude distribution of the events, and then multiplies them together to yield a nearest neighbor distance.

Schuster Test

In order to confirm whether all swarms were being identified in the catalog by both methods of analysis, I applied a Schuster test to the catalog after removing swarms isolated by each method. A Schuster test, first created by Arthur Schuster in 1897, computes the probability that the timing of events in a catalog varies according to a sine-wave function of period T . The probability that the distribution of event times arises from a uniform seismicity rate is referred to as the Schuster p-value. The lower this p-value, the higher the probability that the distribution of the timings of events stacked over the period T is non-uniform, which is usually interpreted as the probability of a periodicity at period T (Ader and Avouac 2013). Therefore, low Schuster p-

values indicate non-uniformities in the catalog, which, in our case, point to swarm events still remaining that have not been identified by the method involved. Once the Schuster p-values have been generated, they can be plotted (Figure 5) to demonstrate whether the observed periodicities at periods T exceed expected values at 99% confidence. Larger numbers of significant periodicities can be interpreted to indicate a greater number of swarm events still remaining in the catalog.

Analysis of Aftershocks and Seismic Swarms

For the analysis of seismic swarms within the catalogs, I used a number of methods established by Jeff Gay, VS2006 and Z2008 in order to test whether to reject or not reject my hypotheses. To identify whether any aftershocks were being identified in the catalog, I used a GUI written by Jeff Gay to manually identify clusters as either swarm-like or aftershock-like, as well as a magnitude difference calculation, which takes the largest magnitude and second largest magnitude event in each cluster of events in the catalog and computes the difference between them.

Next, to identify whether swarms were exponentially decaying in time, I used a time ratio, which divides each cluster into 4 sections of time, counts how many events occur in each quarter, and finally calculates the ratio of number of events in the first quarter to the number of events in the last quarter. If the time ratio is about one, then the swarms do not exponentially decay in time.

In order to test whether the swarms I identified exhibited an even distribution in space, I used a planarity ratio, which calculates the covariance matrix of the demeaned Cartesian coordinates of each event in each swarm cluster, and from that produces three eigenvalues which describe whether the swarm exhibits a linear, planar or spherical distribution. If the swarm is classified as planar, then it demonstrates an even distribution in space.

Finally, to test if a larger fractions of the swarms identified demonstrate migration behavior, I used an expansion ratio, which takes the median distance in each cluster, and then takes the median distance of either half of that swarm, finds the number of events in either half, then finally computes the ratio of the number of events in both medians. If the ratio is large, that indicates the swarm migrated more in the second half than the first, and vice versa. However, if the ratio is about 1, then the swarm migrates fairly evenly in time.

Evaluating Uncertainties in Earthquake Location

The earthquake events in the catalogs I will be working with have been relocated, as previously stated, using the hypoDD, or hypocenter double difference algorithm. This highly precise algorithm results in high-resolution relative hypocenter locations over a large area for each earthquake event, reducing uncertainty by over an order of magnitude compared to catalog locations (Walhauser and Ellsworth, 2000). Although error estimated assigned to relocated hypocenters need to be reviewed, especially when station distribution is sparse or if azimuthal coverage of available phases is not optimal (Walldhauser 2001), Walldhauser and Ellsworth (2000) reviewed a number of error estimates using a battery of tests with the hypoDD algorithm. They were able to conclude that the relocation method is able to image very fine-scale structure of seismicity along fault zones. In addition, they were able to safely conclude that the algorithm allowed for the consistent relocation of seismicity with high resolution along entire fault systems,

therefore corroborating its efficiency and efficacy. Using these conclusions, the vertical relative location error in kilometers at the 95% confidence level is reported as part of Waldhauser's earthquake catalog of Northern California, so those location errors will serve as my uncertainties for the vertical locations of events. Hauksson, Yang and Shearer report similar location error confidence level values for the absolute horizontal position and depth for their events.

Budgetary and Work Plan Considerations

Taken into consideration that the schedule I had previously proposed was somewhat difficult to estimate, as the identification of each swarm can take a variable amount of time, I stayed on track to complete compiling and analyzing my catalogs in California. I had the Northern California catalogs compiled by mid-January, and the Southern California catalogs compiled shortly after. Writing and running the code to be able to analyze the catalogs took through mid-March, and actual analysis was completed by the beginning of April.

The materials I used for my research are as follows:

- 1) Waldhauser's 2013 hypoDD earthquake catalog of Northern California
- 2) Hauksson, Yang and Shearer's 2011 hypoDD earthquake catalog of Southern California
- 3) MATLAB R2015b to analyze earthquake events, programs, etc.

I have not incurred any costs in pursuance of this research.

4) Results

Clustering Analysis Plots: Zaliapin vs. Vidale and Shearer

I first modified or wrote the MATLAB programs for carrying out earthquake catalog analysis outlined in both of the aforementioned methods. To validate my implementation of the Z2008 method, I plotted the $\log(T)$ - $\log(R)$ scatter plots for nearest neighbor earthquakes in the Northern California catalog. Figure 6 shows that the clustering in time and space (Panel a) and bimodal distribution of nearest neighbor distances (Panel b) previously noted by Zaliapin and collaborators is, indeed, a characteristic of the Waldhauser's Northern California catalog. Figure 7 corroborates both of these observations in Southern California, for comparison. Therefore, my implementation of the Z2008 method was indeed capable of identifying clustered distributions of quakes. In order to be able to compare these results to the seismic swarms identified by the VS2006 method, I separated this bimodal distribution into two groups – swarm and non-swarm events – with a cutoff between the clustered and non-clustered distributions at $\eta = -6$.

Next, I applied the VS2006 approach, using the same threshold values as used in their study, in order to identify swarms in Northern California. Then, I applied the Zaliapin clustering analysis to the earthquakes identified as swarms with the VS2006 approach, in order to see where they fell on the $\log(T)$ - $\log(R)$ and nearest neighbor distance plots (Figure 8) in Northern California only, as a validation of methods. Despite the fact that the clustered events distribution is similar for both methods, and the swarm events fall exactly where we expect them to in the $\log(T)$ - $\log(R)$ diagram, it is evident from the histograms in Figure 7b that the VS2006 method only identifies a very small subset of all swarm events.

To analyze the spatial distribution of swarm seismicity identified using the two methods, I plotted them on a map using a program that reads topography and bathymetry data from the Sandwell Database (Sandwell et al. 2009) and plots a topography section (Figure 9) using a set of specified coordinates to produce the desired figure. To confirm that more events were being identified by the former than the latter, I used a program that reads bathymetry data from the Sandwell Database and plots a topography section (Figure 10) using a set of specified coordinates to produce the desired figure. Running this program, I plotted a topography map of Northern California using the latitude limits of 35 to 42 N and longitude limits of -117 to -127 W. I then plotted the seismic swarm events identified by the Z2008 and VS2006 methods in three sets of plots. The swarm events identified by the Z2008 method totaled 21,248 out of a possible 63,705 for a magnitude of completeness $m_c > 2.0$, or about 33% of the catalog. The swarm events identified by VS2006 totaled 4,684 out of 63,705, or about 7.3% of the catalog.

Noting this difference, I then moved on to analyze whether all of the swarm events were being identified by the respective programs using a Schuster Test. The results of this analysis are plotted in Figure 11. I find that a strongly statistically significant periodicity is present in the Schuster spectrum for the catalog with VS2006-identified events removed; this indicates that not all of the seismic swarms present in the catalog are being identified by the approach. In contrast, removing the swarms identified with the Z2008 approach results in no significant periodicities being present in the Schuster spectrum. I take this as an indication that the Z2008 method identifies all the swarms.

Having confirmed that the choice of thresholds used by VS2006 did not identify all the seismic swarms present in Waldhauser's Northern California catalog (from here out referred to as NCA), I proceeded to analyze the effects of varying each of the five threshold values. The threshold number of days was varied between 10 and 48, minimum number of earthquakes between 21 and 59, the maximum number of earthquakes preceding a swarm sequence between 0 and 17, and the minimum and maximum distances between 1 and 15 and 0 and 18, respectively. These variations on the threshold values resulted in a total of 5 different swarm catalogs based on the VS2006 approach, which I then compared against the swarm events identified by the Z2008 analysis. I noted how adjusting each threshold parameter affected the number of swarm events identified.

In order to compare the adjusted V&S thresholds to the clustering analysis, I processed the swarm events identified by the adjusted V&S thresholds to find the nearest neighbor distance given by the clustering analysis, and then plotted all five of the nearest neighbor distance distributions given by the adjusted thresholds against the clustering analysis nearest neighbor distance distribution (Figure 12). I used the first 50,000 earthquakes to examine these distributions due to time constraints associated with running the analyses on the entire catalog for each threshold adjustment.

For the time threshold, an increase in the number of days resulted in the distribution becoming much more spread out in space without much increase in the number of events identified. This was the same for the N2 parameter, or the threshold that describes the number of earthquakes (at least 39) after the initial event in the swarm sequence). Increasing the N1 threshold, or the number of earthquakes preceding the initiating event (3 or fewer), resulted in a slight increase in the number of events identified. The minimum and maximum distance

thresholds, however, had the largest effect on the number of events identified. As these distances were increased, there was a dramatic increase in the number of events identified and an increase in similarity to the swarm distribution inferred by the Z2008 analysis.

Having completed this analysis, I proceeded to plot the distribution of swarm and non-swarm events in time, depth, and magnitude in order to investigate how the events identified by both methods compared based on these parameters (Figures 11, 12, 13).

In all three of the different plotted distributions, it was evident that while both methods were identifying the same sets of swarms, the VS2006 approach was not identifying all the events constituent of those swarms, whereas the Z2008 method seemed to be. To investigate why, I decided to manually identify discrepancies between the two plots and identify which swarms the Z2008 approach had identified but that VS2006 had missed.

Extending the Original Thresholds and Catalog Analysis

Overall, there were 85 clusters of events that Z2008 had identified, but that VS2006 had not. Once I had these clusters documented and characterized based on the number of events within the cluster temporal and spatial distribution as well as where they occurred, I went through and analyzed what thresholds each individual cluster was missing that would have it be missed by VS2006, but identified by Z2008. One example of an excluded swarm is demonstrated in Figures 14. It occurred in 1993 between 05/17-05/29 (9 days in total) near Mt. Whitney in the Long Valley Caldera region of Northern California (also the site of the 1986 Mammoth Mountain swarm, which has been rigorously studied and documented in literature) and contains 137 events. Although these numbers are consistent with VS2006, the swarm occurs over a distance of 39.53 km, which is over the allowed threshold given by the analysis. Therefore, although this cluster is definitely a swarm, the thresholds would miss it based on distance constraints. I found that through my analysis overall, the greatest controls on whether a swarm would be identified by the thresholds was the distance it was allowed to occur in (D1 and D2), the minimum number of events that could be defined as a swarm (N1), and the number of days it was allowed to occur over (T). Furthermore, I found that by extending T, the number of earthquakes allowed to occur in that prior number of days within the same radius of the swarm (N2) had to expand as well. Once I had this established, I went back through and found the minimum and maximum T, N1, N2, D1 and D2 that characterized all of the clusters of events that had been missed by VS2006, and used these to extend the original threshold parameters. I first compared my adjusted thresholds to the original thresholds using Z2008 and plotting a comparison of the thresholds on a $\log(T)$ - $\log(R)$ plot as well as a nearest neighbor distribution (Figures 15 and 16). The comparison of plots and distributions between the original and adjusted thresholds yielded that while the adjusted thresholds did tend to identify more earthquakes from the non-clustered distribution, they still identified far more events for both Northern and Southern California associated with the clustered events than the original thresholds.

As is evident from the comparison of the time distributions between the original and adjusted thresholds (Figure 17), not only do the adjusted thresholds match the number of events consistent of clusters identified by Z2008, but actually identify more events than the clustering analysis, indicating that adjusting and extending the threshold values is not only more complete than the original thresholds, but might potentially identify more swarms than Z2008. In addition,

by plotting the comparison of the original and adjusted thresholds on a map of Northern California (Figure 18), it is clear that by extending the thresholds, the clusters identified become less isolated bursts of seismic activity and longer, more linear trends of activity along seismically active areas, which indicates that adjusting the thresholds control how they identify events in clusters. Overall, the original thresholds identified a total of 56 clusters of events, while the adjusted thresholds identified 1339 clusters of events, 9-fold what the original thresholds identified. With this in mind, I also tested the original and adjusted threshold values for a time and map distribution in Southern California (Figures 19 and 20), due to the fact that VS2006 was originally established and tested using data from Southern California. As Southern California is characterized as an extensional setting, whereas Northern California is characterized as a strike-slip setting, it was important to test the thresholds for both areas to make sure that there wasn't a statistically significant difference in how the thresholds identified swarms in both areas. However, the time and map distributions for Southern California demonstrate, consistent with the results of Northern California, that the adjusted thresholds identify more seismic bursts and more events associated with the bursts than the original thresholds. The original thresholds identified 131 clusters, while the adjusted thresholds identified 1868 clusters, 14-fold of what the original thresholds identified. Therefore, I could safely establish that my adjusted thresholds resulted in a more complete swarm catalog than the original thresholds established by VS2006.

Merging the Catalog and Removing Aftershocks

After I was able to establish that the adjusted threshold values were more effective at identifying clusters of seismicity in both Northern and Southern California than the original thresholds, I needed to identify any aftershock sequences present in the catalogs. Aftershock sequences are characterized by a magnitude difference between the first largest and second largest event of 1.1-1.2, as described by Omori's law, as well as an exponential decay in the number of events with time. As one of my hypotheses is that seismic swarms will not exhibit an exponential decay in the number of events with time, having aftershocks present in the catalog would skew my results. Therefore, it was necessary to be able to effectively identify any aftershocks present in the catalogs and remove them before being able to continue on with my analysis. I first used the Swarm Categorizer to manually pick through and identify swarm and aftershock sequences for both Northern and Southern California, then applied the magnitude difference calculation as a check on my results to account for human error. I found that for Northern California, using the Swarm Categorizer identified 58.93% of the original thresholds catalog to be aftershocks. In comparison, only 27.01% of the adjusted thresholds catalog was aftershocks. In Southern California, however, the results were much lower and more similar: 16.26% of the original thresholds catalog was aftershocks, while 14.09% of the adjusted thresholds catalog was swarms. For comparison, the magnitude difference computation yielded similar results for both regions: In Northern California, 8.93% and 5.22% aftershocks, respectively, and in Southern California, 11.45% and 7.12%, respectively. Therefore, there were aftershocks present in both catalogs that needed to be manually scrubbed. I further proceeded by also plotting the start, end and mean locations of all clusters for both Northern and Southern California on a map to see where they occurred and if there was any evidence of spatial migration (Figures 21 and 22). What is immediately apparent is that some of the clusters appear to be related to one another based on how they migrate and what their position is in space. In addition, it is clear that a higher fraction of clusters in appear to migrate in in the adjusted thresholds catalog compared to the original thresholds catalog. Therefore, the other method I

needed to apply to manually prep to catalog was to combine related clusters so as to not miss any aftershocks that were “chopped up” by the thresholds and could be mistaken as a swarm in analysis, as well as make sure not to misrepresent the number of swarms being identified.

Analysis of Catalogs

Once I had manually “scrubbed” the original and adjusted catalogs for both merged clusters and for aftershocks, I found that the number of clusters identified by the thresholds decreased in both number and in position in space (Figures 23 and 24). The original threshold catalogs in Northern and Southern California decreased from 56 to 37 swarms, and 131 to 108 swarms, respectively, with a total number of events identified as 4499 and 12,803, respectively. The adjusted threshold catalogs also decreased from 1339 to 1240 swarms, and from 1868 to 1730 swarms, respectively, with an overall number of events identified as 39,756 and 83,479 events, respectively. Consequently, although the adjusted threshold catalogs experienced a greater decrease in the number of swarms identified, they still contain more swarms and more swarm events overall than either of the original threshold catalogs for Northern or Southern California. Next, I was able to go through and classify each catalog based on the magnitude difference, expansion ratio and planarity distribution.

Calculating the time ratio exhibited consistent results for the original and adjusted threshold catalogs. The original thresholds in Northern and Southern California demonstrated that 1/37 and 10/108 swarms or 2.7% and 9.3%, respectively, had a time ratio between 0.8-1.5, indicating that they did not exponentially decay in time. The adjusted thresholds demonstrated that 323/39,756 and 376/83,479 swarms, or 0.81% and 0.45%, respectively had a time ratio between 0.8-1.5, which indicates that although a lower percentage of swarms did not exponentially decay in the adjusted thresholds compared to the original, the adjusted thresholds still identified a larger number of swarms overall. For the expansion ratio, only 9 swarms out of a total of 37, or 24.3% for the NCA original thresholds exhibited an expansion ratio between 0.8-1.5, indicating the swarms had an even distribution in space. This was similar for the SCA original thresholds, where only 26 out of a total of 108, or 24.07% exhibited an expansion ratio between 0.8-1.5. Conversely, for the adjusted thresholds this number was far higher, where 455/39,756, or 1.14% were shown to migrate evenly in NCA, and 1730/83,479, or 2.04% of swarms in SCA. Hence, despite the fact that a smaller fraction of the swarms identified were shown to migrate for the adjusted thresholds, a larger number of swarms overall were both identified and shown to migrate compared to the original thresholds. Finally, for the planarity distribution computation, the results were a little more divided: 3/37, or 8.01% and 94/108, or 87.04% for the original thresholds in NCA and SCA, respectively exhibited a planarity distribution between 0.8-1.5, indicating that the swarms were migrating. For the adjusted thresholds, 1052/39756 or 2.06% and 1456/83479, or 1.7% of swarms in NCA and SCA, respectively, were shown to migrate. Thus, a larger number of swarms were both shown to migrate and had a more even distribution in space in the adjusted thresholds catalogs compared to the original thresholds.

5) Discussion and Conclusions

VS2006 vs Z2008

It is evident from my analysis and results thus far that although both methods are effective in identifying seismic swarms, the Z2008 clustering analysis method is more effective in identifying not only the sequences which swarm events belong to, but a higher percentage of the events than the arbitrary thresholds proposed by VS2006. However, the VS2006 approach is more effective in only identifying seismic swarm sequences without mistakenly involving other, unrelated clusters. This is in contrast to the Z2008 method, which identifies more earthquakes related to one another in space and time without regard to which swarm those events belong. It is interesting to note that although the two methods identified very different numbers of total swarm events, neither program seemed to be affected by the background seismicity present in the Wauldhauser catalog. In addition, it is also interesting to note that from the time distribution, the VS2006 approach identified no events that occurred in a few years, including 1985 and 1989. Upon further analysis, any swarms missed in these years did not meet the minimum number of events (39) required by the original thresholds, which constituted a major control on which clusters would be identified by the thresholds.

Original vs. Adjusted Thresholds

By expanding and adjusting the original thresholds to include clusters missed by the thresholds, both more clusters of events and more events were identified overall in Northern and Southern California than by either the original thresholds or by the Z2008 clustering analysis. Based on the $\log(T)$ - $\log(R)$ and nearest neighbor distance distribution analysis alone, it appeared that the adjusted thresholds were picking up more non-swarm events than the original thresholds, which would lead to skewed results. However, because manual scrubbing of the catalogs to merge clusters and remove aftershocks was required, this concern was not relevant to the investigation and the adjusted thresholds still yielded identification of more clustered events overall. Analysis of these catalogs once related clusters were merged and aftershocks were removed yielded the results that the adjusted thresholds did demonstrate that swarms did not exponentially decay with the number of events in time, that a larger fraction of swarms exhibited a more even distribution in space than found by VS2006, and had a larger fraction shown to migrate than was found by Chen and Shearer (2011). However, it is interesting to note that the original thresholds in Southern California identified over 80% of the swarms contained in the catalog to migrate in space, which could indicate that for Southern California based on its extension setting, demonstrates a difference in how swarms migrate and how the thresholds work to identify swarms in this tectonic regime.

Looking at the time ratios that didn't indicate swarms that had an even distribution in time that the original and adjusted catalogs were more related to one another than initial thought. The original catalogs for both Northern and Southern California both demonstrated that most of the swarms had a time ratio much larger than 1 (with a mean time ratio of 10.05 and 11.7, respectively), indicating that more events occurred in the first quadrant than the last. In comparison, swarms in the adjusted catalogs for Northern and Southern California had a mean time much closer to 1 (2.1 and

2.3, respectively), indicating that although swarms still did have a larger number of events in the first quadrant, the difference wasn't nearly as significant.

Analysis of plotting the magnitude difference as a histogram distribution as well as against the expansion ratio and the counts ratio (defined as the number of the events in the last quadrant of time of each swarm divided by the number of events in the first quadrant of time) yielded some interesting results. The original thresholds for NCA and SCA demonstrated no clear trend or distinctive function that would describe the distribution for the magnitude difference vs expansion ratio (Figure 25). In addition, the original thresholds exhibited a clustering of most swarms towards the beginning of the plot, rather than being distributed more evenly throughout. The original threshold catalog for NCA also demonstrated a bimodal distribution for the frequency of magnitude difference, indicating there could be two statistically distinct sets of swarms in Northern California based on its strike-slip tectonic setting.

However, the adjusted threshold catalog for NCA demonstrated completely different results. For the expansion ratio vs. magnitude difference (Figure 26), there was a definite exponential decay. Taken into consideration with the results of the planarity distribution and expansion ratio, I would not interpret this to mean that the adjusted thresholds are not finding swarms or swarms that exponentially decay, but rather that so many events are being identified that the plots demonstrate a complete swarm regime, compared to the original thresholds in NCA, which do not demonstrate a complete catalog.

6) Summary

In summary, I have used two different methods to identify seismic swarms from Vidale and Shearer (2006) and Zaliapin et al. (2008), respectively and used them on earthquake catalogs identified by Waldhauser (2013) in Northern California and Hakusson, Yang and Shearer (2013) in Southern California in order to identify sequences of swarm events in the aim of seeing which method more effectively identified swarm events. I have concluded from analysis of timing and frequency of events that the Zaliapin et al. method more effectively identifies swarm events in comparison to Vidale and Shearer.

From that, I adjusted the original thresholds identified by VS2006 and created my own swarm catalog using those thresholds, then compared the original and adjusted catalogs produced by both respective thresholds. I analyzed all swarm sequences contained within the catalogs and concluded that the original swarm thresholds did not identify all seismic swarms in either Northern or Southern California. In addition, I was able to accept my hypotheses and demonstrate that the swarms I identified using my adjusted thresholds demonstrated that a larger number did not exponentially decay, had a more even distribution in space, and finally a larger fraction migrated in time.

Bibliography

- [1] Ader, T. J., and J.-P. Avouac. Detecting periodicities and declustering in earthquake catalogs using the Schuster spectrum, application to Himalayan seismicity, *Earth Planet. Sci. Lett.*, 377–378, 97–105, 2013.
- [2] Baiesi, Marco, and Maya Paczuski. "Scale-free Networks of Earthquakes and Aftershocks." *Physical Review E Phys. Rev. E* 69.6, 2004.
- [3] Chen, X., and P. M. Shearer. "Comprehensive Analysis of Earthquake Source Spectra and Swarms in the Salton Trough, California." *J. Geophys. Res. Journal of Geophysical Research* 116.B9, 2011.
- [4] Chouliaras, G., G. Drakatos, K. Pavlou, and K. Makropoulos. "Stress Distribution and Seismicity Patterns of the 2011 Seismic Swarm in the Messinia Basin, (South-Western Peloponnesus), Greece." *Nat. Hazards Earth Syst. Sci. Natural Hazards and Earth System Science* 13.1 (2013): 45-51.
- [5] Corral, Álvaro. "Local Distributions and Rate Fluctuations in a Unified Scaling Law for Earthquakes." *Physical Review E Phys. Rev. E* 68.3, 2003. [1] Holtkamp, S.g., and M.r. Brudzinski. "Earthquake Swarms in Circum-Pacific Subduction Zones." *Earth and Planetary Science Letters* 305.1-2 (2011): 215-25.
- [6] Fischer, Tomáš, and Josef Horálek. "Space-time Distribution of Earthquake Swarms in the Principal Focal Zone of the NW Bohemia/Vogtland Seismoactive Region: Period 1985–2001." *Journal of Geodynamics* January-March 35.1-2 (2003): 125-44.
- [7] Hauksson, E. and W. Yang, and P.M. Shearer, "Waveform Relocated Earthquake Catalog for Southern California (1981 to 2011)"; *Bull. Seismol. Soc. Am.*, Vol. 102, No. 5, pp.2239-2244, October 2012, doi:10.1785/0120120010.
- [8] Lohman, R. B., and J. J. Mcguire. "Earthquake Swarms Driven by Aseismic Creep in the Salton Trough, California." *J. Geophys. Res. Journal of Geophysical Research: Solid Earth* 112.B4 (2007)
- [9] Parotidis, M., E. Rothert, and S. A. Shapiro. "Pore-pressure Diffusion: A Possible Triggering Mechanism for the Earthquake Swarms 2000 in Vogtland/NW-Bohemia, Central Europe." *Geophys. Res. Lett. Geophysical Research Letters* 30.20 (2003).
- [10] Ruppert, N. A., S. Prejean, and R. A. Hansen. Seismic swarm associated with the 2008 eruption of Kasatochi Volcano, Alaska: Earthquake locations and source parameters, *J. Geophys. Res.*, 116, 2011.
- [11] Sandwell, David T., and Walter H. F. Smith. "Global Marine Gravity from Retracked Geosat and ERS-1 Altimetry: Ridge Segmentation versus Spreading Rate." *J. Geophys. Res. Journal of Geophysical Research* 114.B1 (2009).
- [12] Savage, M. K., V. Ferrazzini, A. Peltier, E. Rivemale, J. Mayor, A. Schmid, F. Brenguier, F. Massin, J.-L. Got, J. Battaglia, A. Dimuro, T. Staudacher, D. Rivet, B. Taisne, and A. Shelley. "Seismic Anisotropy and Its Precursory Change before Eruptions at Piton De La Fournaise Volcano, La Réunion." *J. Geophys. Res. Solid Earth Journal of Geophysical Research: Solid Earth* 120.5 (2015): 3430-458.
- [13] Shapiro, S.a., and C. Dinske. "Fluid-induced Seismicity: Pressure Diffusion and Hydraulic Fracturing." *Geophysical Prospecting* 57.2 (2009): 301-10.
- [14] "Schuster Spectrum – Searching for Periodicities in the timing of Earthquakes." California Institute of Technology Tectonics Observatory, 5 Oct. 2012.
- [15] Thatcher, W., and J. N. Brune. "Seismic Study of an Oceanic Ridge Earthquake Swarm in the Gulf of California." *Geophysical Journal International* 22.5 (1971): 473-89.

- [16] Umeda, Koji, Koichi Asamori, Ayumu Makuuchi, Kazuo Kobori, and Yuki Hama. "Triggering of Earthquake Swarms following the 2011 Tohoku Megathrust Earthquake." *J. Geophys. Res. Solid Earth Journal of Geophysical Research: Solid Earth* 120.4 (2015): 2279-291.
- [17] U.S. Geological Survey and California Geological Survey, 2006, Quaternary fault and fold database for the United States, accessed Nov 12, 2015, from USGS web site: <http://earthquake.usgs.gov/hazards/qfaults/>.
- [18] Vidale, John E., and Peter M. Shearer. "A Survey of 71 Earthquake Bursts across Southern California: Exploring the Role of Pore Fluid Pressure Fluctuations and Aseismic Slip as Drivers." *J. Geophys. Res. Journal of Geophysical Research* 111.B5 (2006).
- [19] Waldhauser, F., HypoDD: A computer program to compute double-difference earthquake locations, USGS Open File Rep., 01-113, 2001.
- [20] Waldhauser F. and W.L. Ellsworth, A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, *Bull. Seism. Soc. Am.*, 90, 1353-1368, 2000.
- [21] Waldhauser, Felix, and David P. Schaff. "Large-scale Relocation of Two Decades of Northern California Seismicity Using Cross-correlation and Double-difference Methods." *J. Geophys. Res. Journal of Geophysical Research* 113.B8 (2008).
- [22] Zaliapin, Ilya, and Yehuda Ben-Zion. "Asymmetric Distribution of Aftershocks on Large Faults in California." *Geophysical Journal International* 185.3 (2011): 1288-304.
- [23] Zaliapin, I., and Y. Ben-Zion. Earthquake clusters in southern California I: Identification and stability, *J. Geophys. Res. Solid Earth*, 118, 2847–2864, 2013.
- [24] Zaliapin, I., and Y. Ben-Zion (2013), Earthquake clusters in southern California II: Classification and relation to physical properties of the crust, *J. Geophys. Res. Solid Earth*, 118, 2865–2877, *doi:10.1002/jgrb.50178*.
- [25] Zhang, Qiong, and Peter M. Shearer. "A New Method to Identify Earthquake Swarms Applied to Seismicity near the San Jacinto Fault, California." *Geophys. J. Int. Geophysical Journal International* 205.2 (2016): 995-1005.

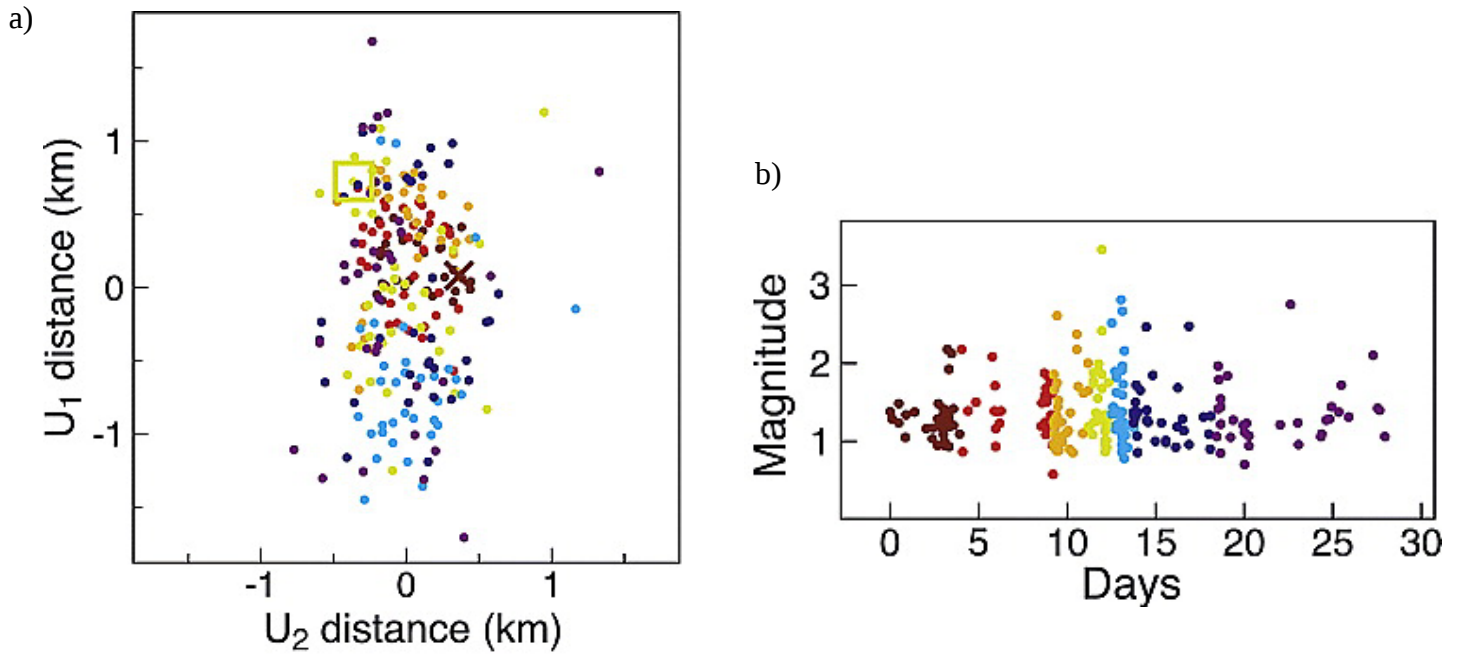


Fig. 1. a) Figure from Vidale and Shearer (2006), illustrating a swarm-like cluster of 230 earthquakes. Colors denote divisions of days. Note the linearity of the morphology of the swarm, which is characteristic of swarm sequences. b) Figure from Vidale and Shearer (2006) demonstrating the distribution of swarm event magnitudes over days. Note that the largest event does not occur at the beginning, but, rather in the middle of the sequence, which is characteristic of swarms.

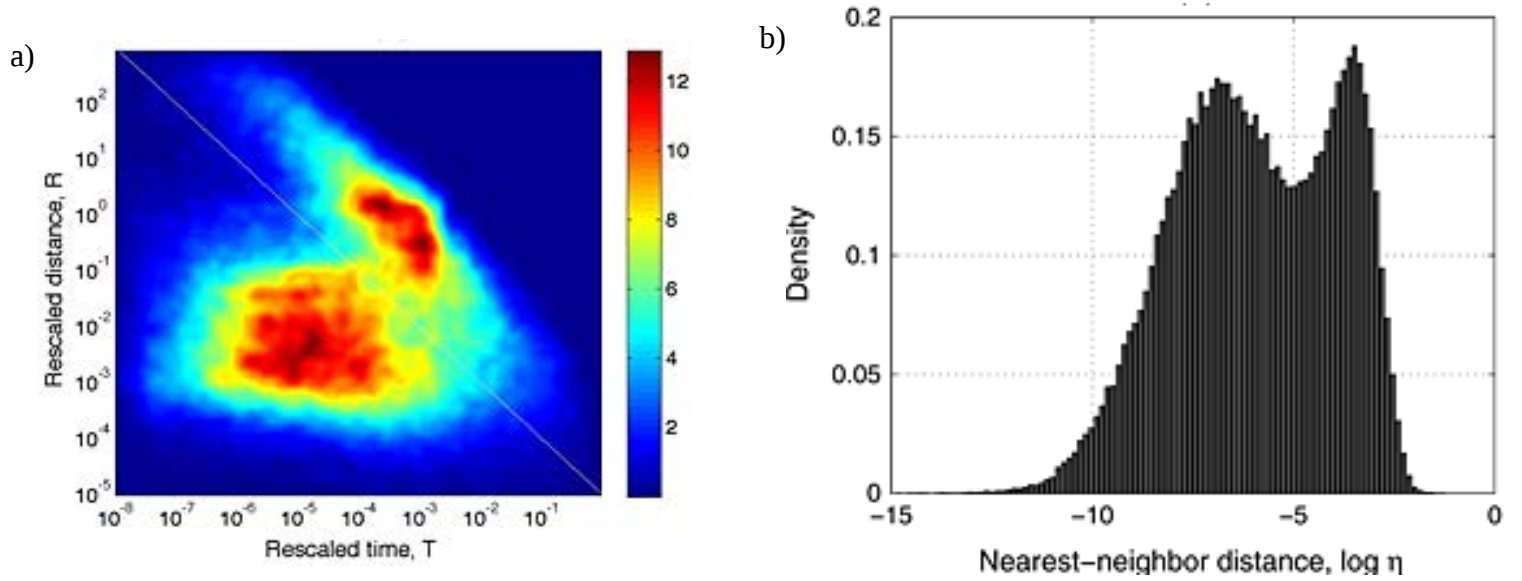
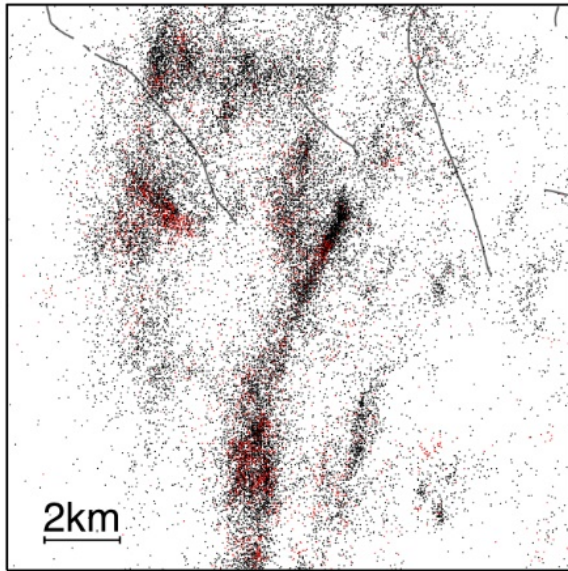


Fig. 2. a) Figure from Zaliapin and Ben-Zion (2013) illustrating the magnitude normalized nearest-neighbor time-distance plot showing a statistically distinct bimodal distribution of the clustered and non-clustered events in space and time. b) Figure from Zaliapin and Ben-Zion (2013) illustrating the bimodal distribution of η that results from the nearest neighbor distance distribution. Clustered (including swarm) events are to the left of -5, non-clustered events to the right.

a)



b)

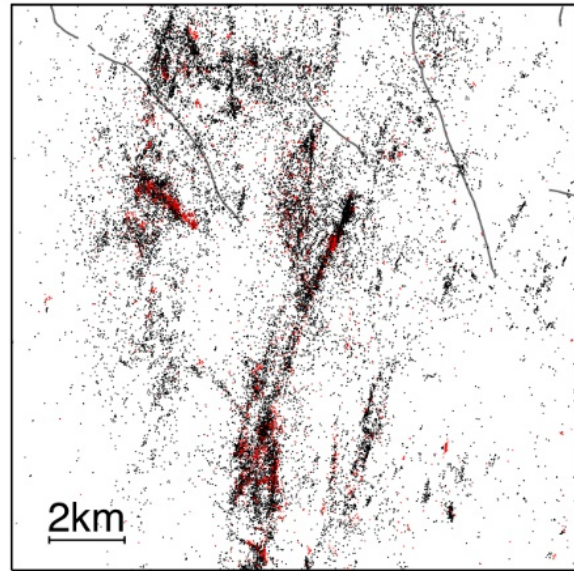


Fig. 3. a) Figure from Waldhauser (2008) demonstrating earthquake locations before application of hypoDD algorithm correction. b) Earthquake locations after application of hypoDD algorithm correction. Note how much more clearly earthquake locations are plotted in space, demonstrating the precision of the correction algorithm.

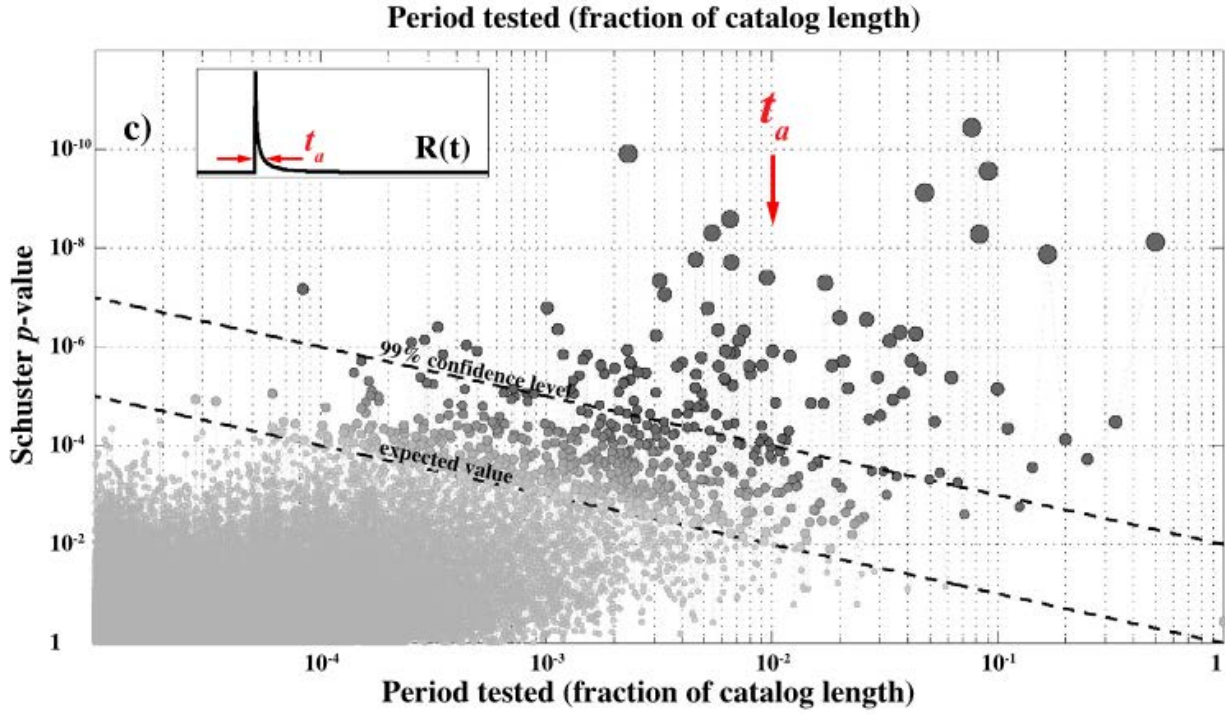


Fig. 4. Schuster plot of an event catalog containing an aftershock sequence, courtesy of Ader and Avouac (2013). Note the number of periodicities that occur above the 99% confidence level. In our analyses, such an anomalously high probability could potentially point to the presence of swarm events in the catalog analyzed.

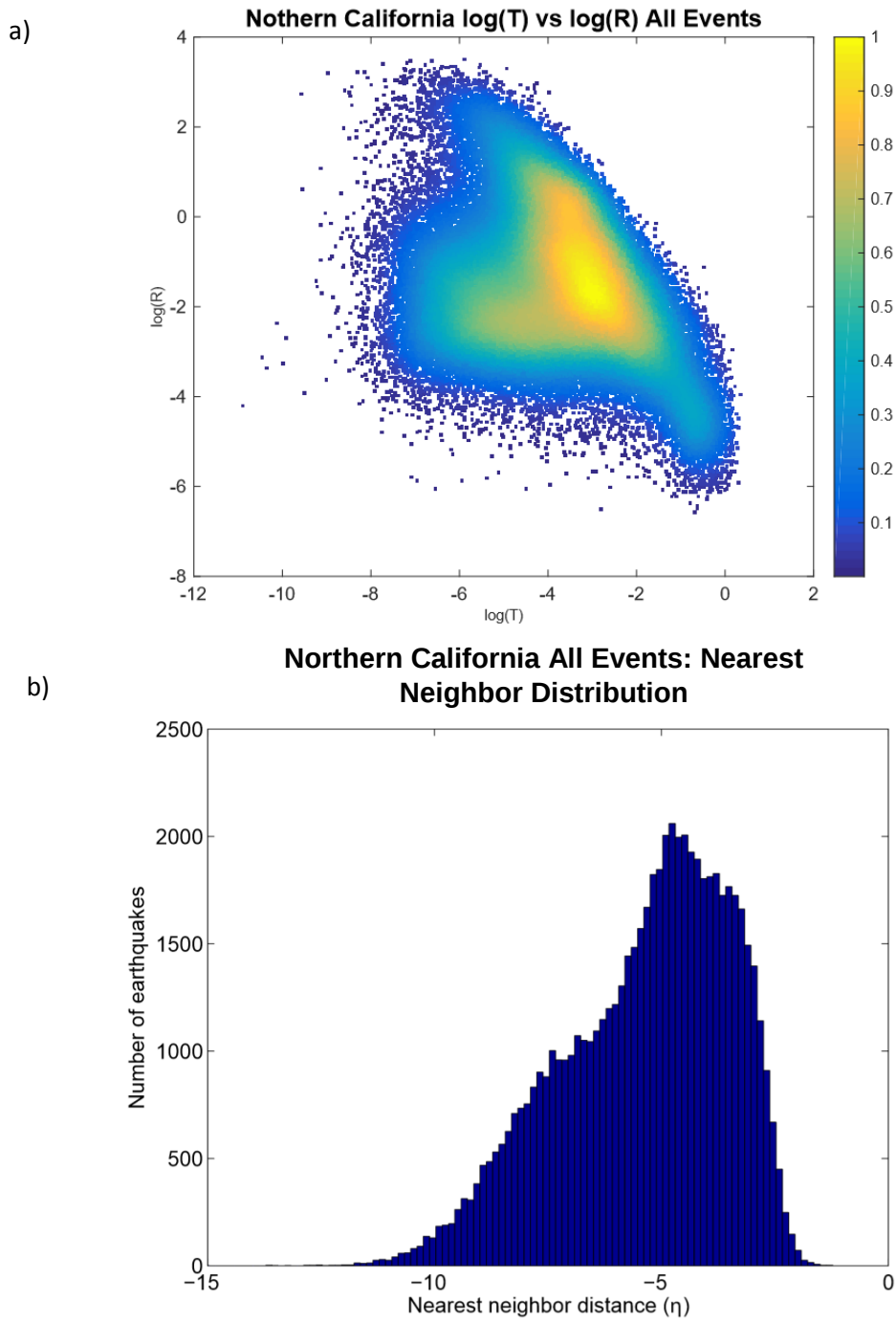


Fig. 5. a) $\log(T)$ - $\log(R)$ plot for Northern California, with a distinct bimodal distribution given by the density distribution. b) Note that two distributions seem to merge at $\sim \eta = -6$, with anomalously clustered events having higher negative value of η .

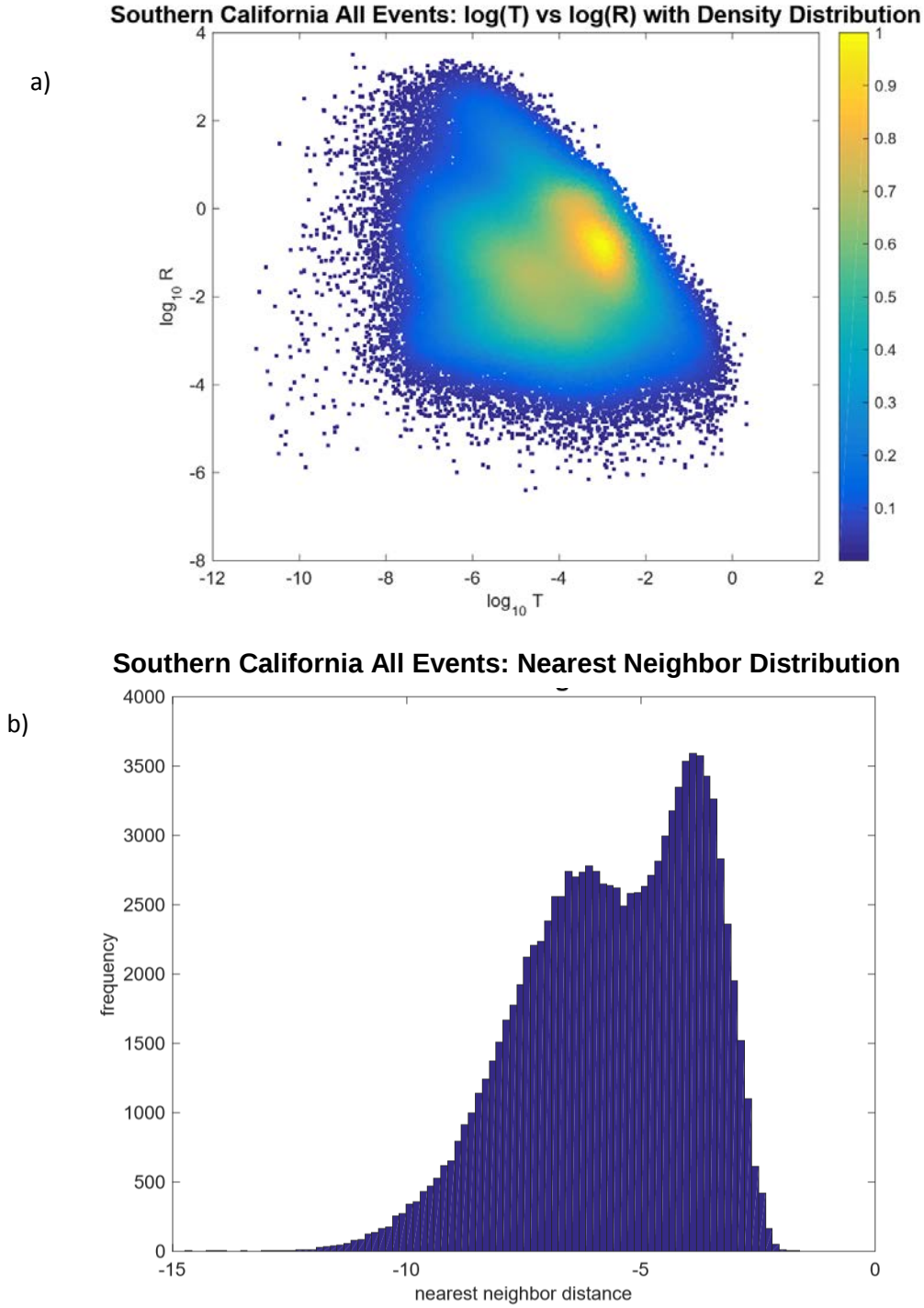
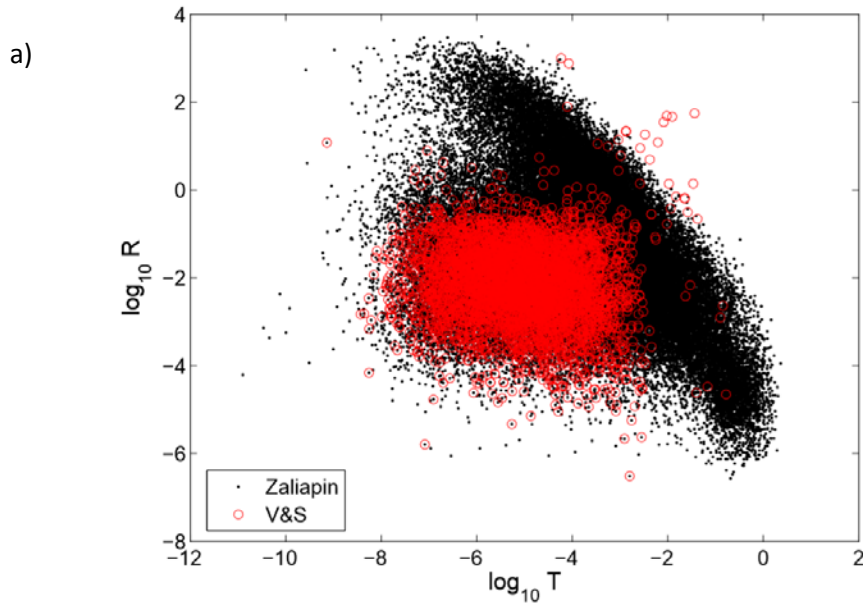


Fig. 6. a) $\log(T)$ - $\log(R)$ plot for Southern California. Note that the density distribution for the events is still separated, similar to Northern California, but less distinctly. b) Nearest neighbor distance distribution for Southern California. Note that two distributions seem to merge at $\sim \eta = -6$, with anomalously clustered events having higher negative value of η , but with a tighter, higher distribution than Northern California exhibits.

Z2008 vs VS2006: $\log(T)$ - $\log(R)$ for all Events in Northern California



Z2008 vs VS2006: Nearest Neighbor Distribution for all Events in Northern California

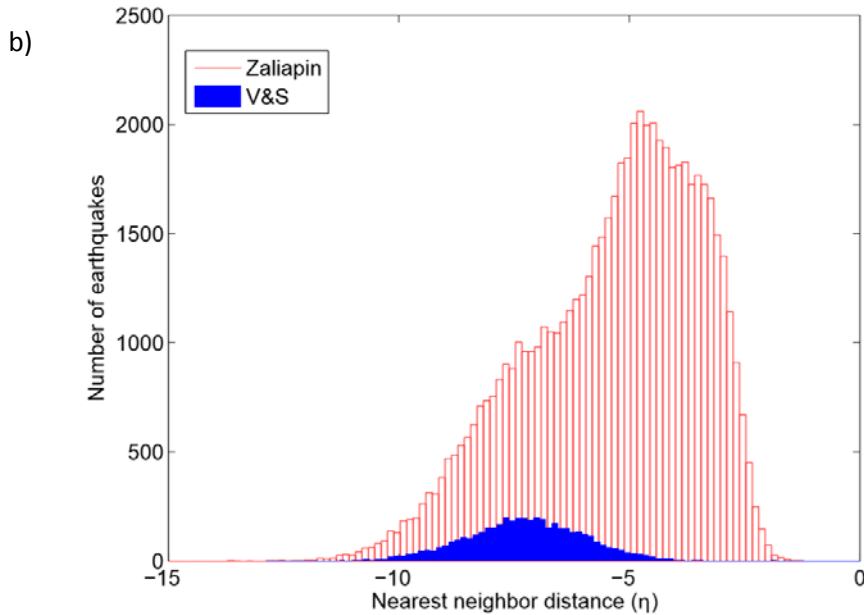
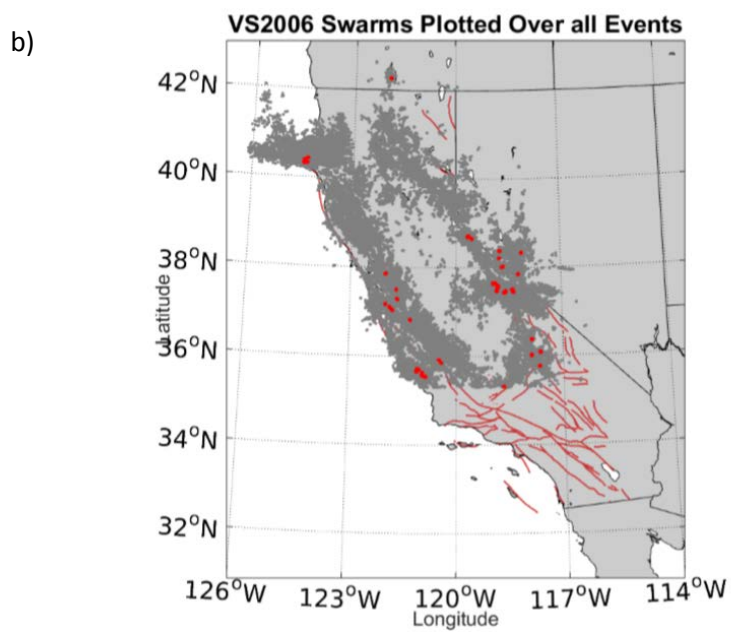
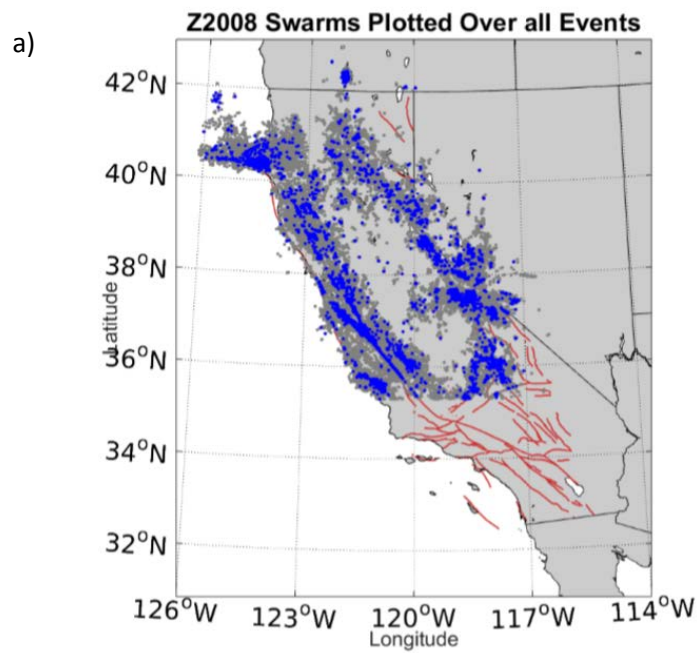


Fig. 7. a) $\log(T)$ - $\log(R)$ plot following the Z2008 analysis for all earthquakes (black) and earthquakes identified as swarm events by the VS2006 (red) method. Note that the swarm events identified using the VS2006 method fall in the region of the plot expected for clustered seismicity. b) Nearest neighbor distance distributions for all events (red) and events identified as swarms by the VS2006 (blue) method. Note that the VS2006 swarm events are much more normally distributed and fewer in number than the clustered part of the distribution.



c)

NCA Original Thresholds: VS2006 and Z2008 Plotted Over All Events

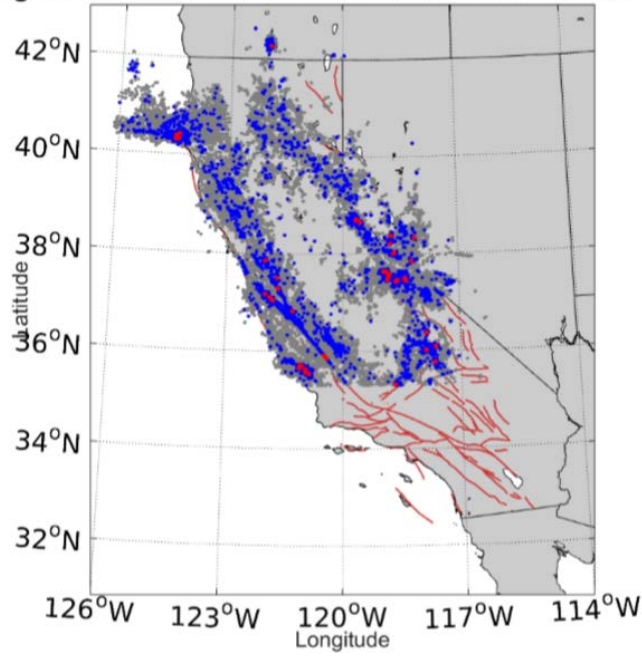


Fig. 8. a) Full map of swarm distributions over Northern California. In grey are all the events contained in Waldhauser's 2013 catalog, in blue are the swarm events identified by Zaliapin. b) Full map of swarm distributions over Northern California. In grey are all the events contained in Waldhauser's 2013 catalog, in red are the swarm events identified by Vidale and Shearer. c) Full map of swarm distributions over Northern California. In grey are all the events contained in Waldhauser's 2013 catalog, in blue are the swarm events identified by Zaliapin, and in red are the events identified by V&S. Note that Z2008 identifies more events with less spatial clustering than VS2006, which focuses on isolated bursts of seismic activity.

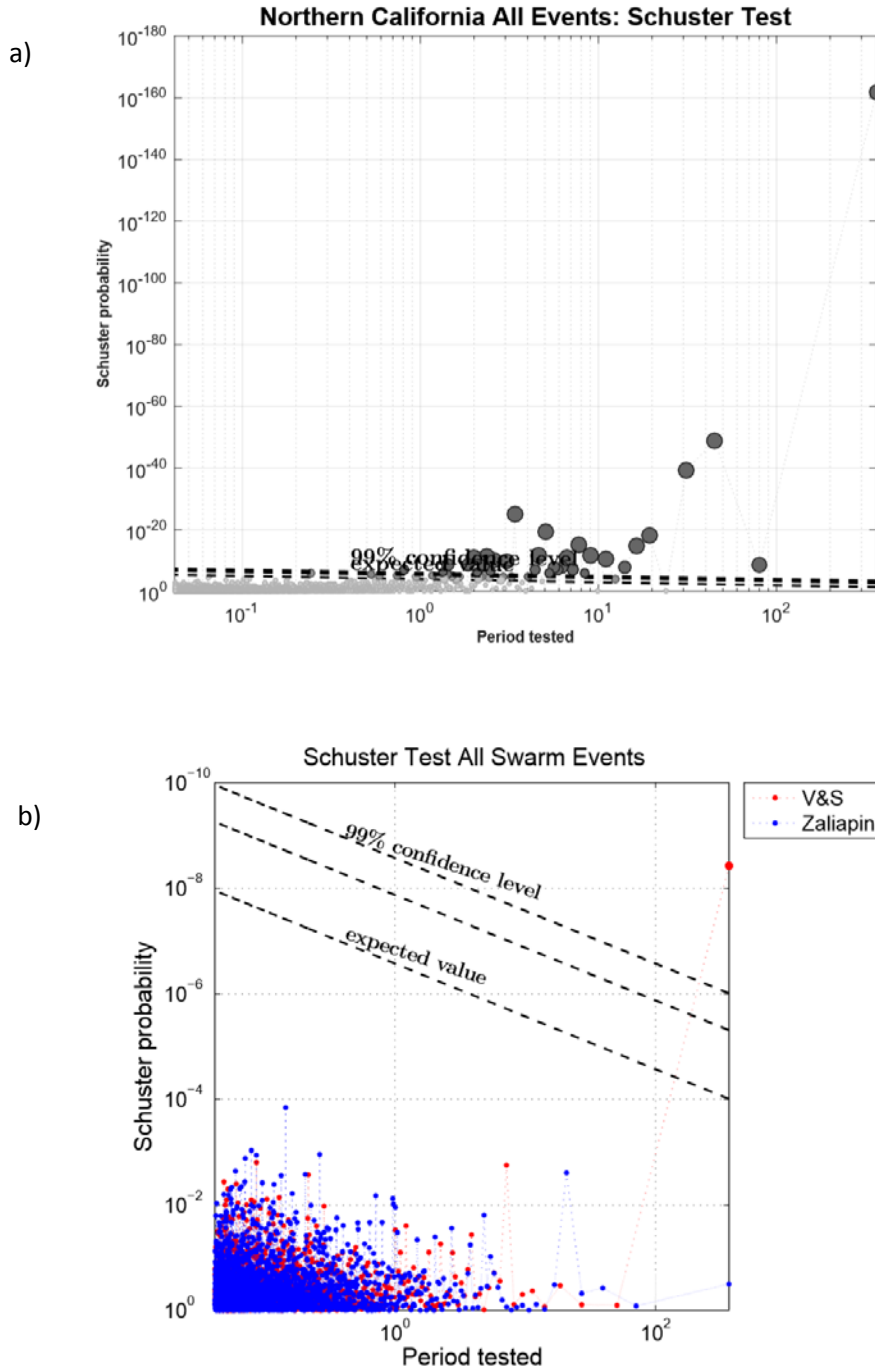


Fig. 9. a) Schuster spectrum for all the events contained in Walhauser's Northern California catalog. Each periodicity above the 99% confidence line indicates an aftershock or anomalously clustered event present in the catalog. b) Schuster spectrum plotted from testing the periodicities present in the data once all swarm events identified by both methods are removed. VS2006 is in red and Z2008 is in blue. Note that the one periodicity in red above the 99% confidence level demonstrates that there are periodicities present in the data, indicating that VS2006 does not identify all swarm events present in Waldhauser's catalog.

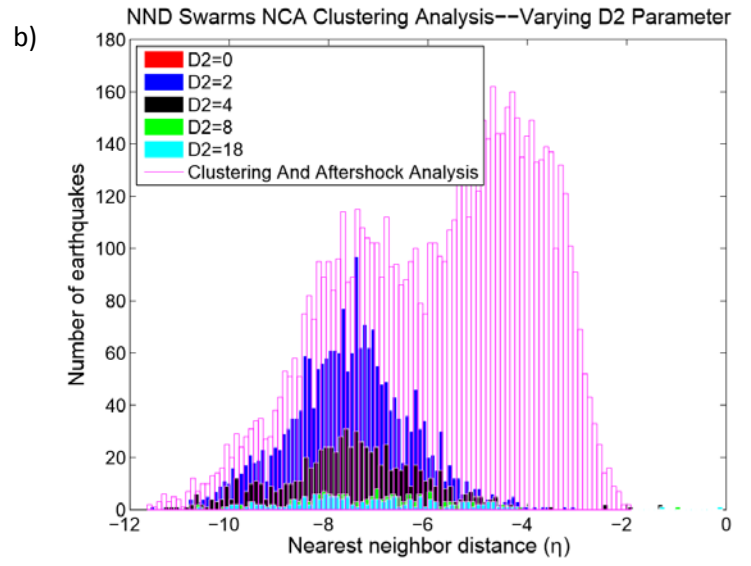
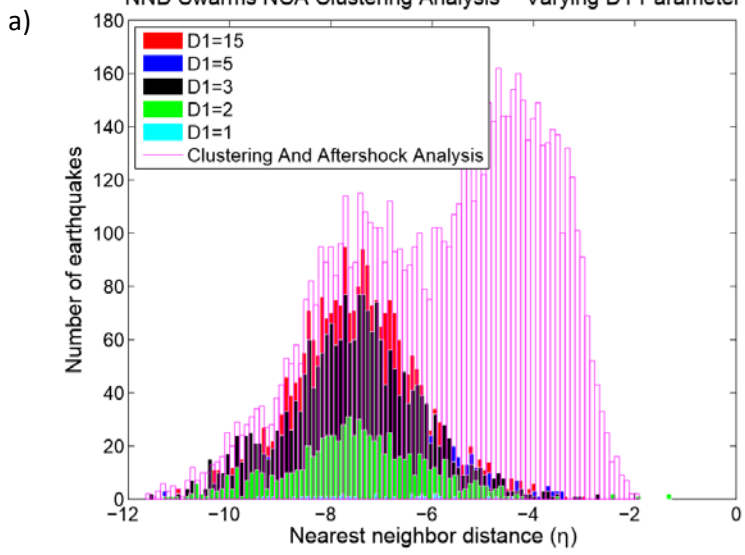


Fig. 10. a) Nearest neighbor distance distribution for threshold variation of the minimum distance parameter. The magenta color corresponds to all the events identified by Z2008. Note that as the minimum distance threshold is increased, the number of events identified increases and begins to match the swarms identified by the clustering analysis. The minimum distance varied is $D1=1$ due to the fact that the minimum distance cannot be 0. b) Nearest neighbor distance distribution for the threshold variation of the maximum distance parameter. The magenta color corresponds to all the events identified by Z2008. Note that as the maximum distance threshold is increased, the number of events identified increases and begins to match the swarms identified by the clustering analysis. Only these two figures are shown because they show the greatest results for the variation of threshold parameters, and only the first 50,000 earthquakes from the Waldhauser catalog were used to make these figures due to time constraints.

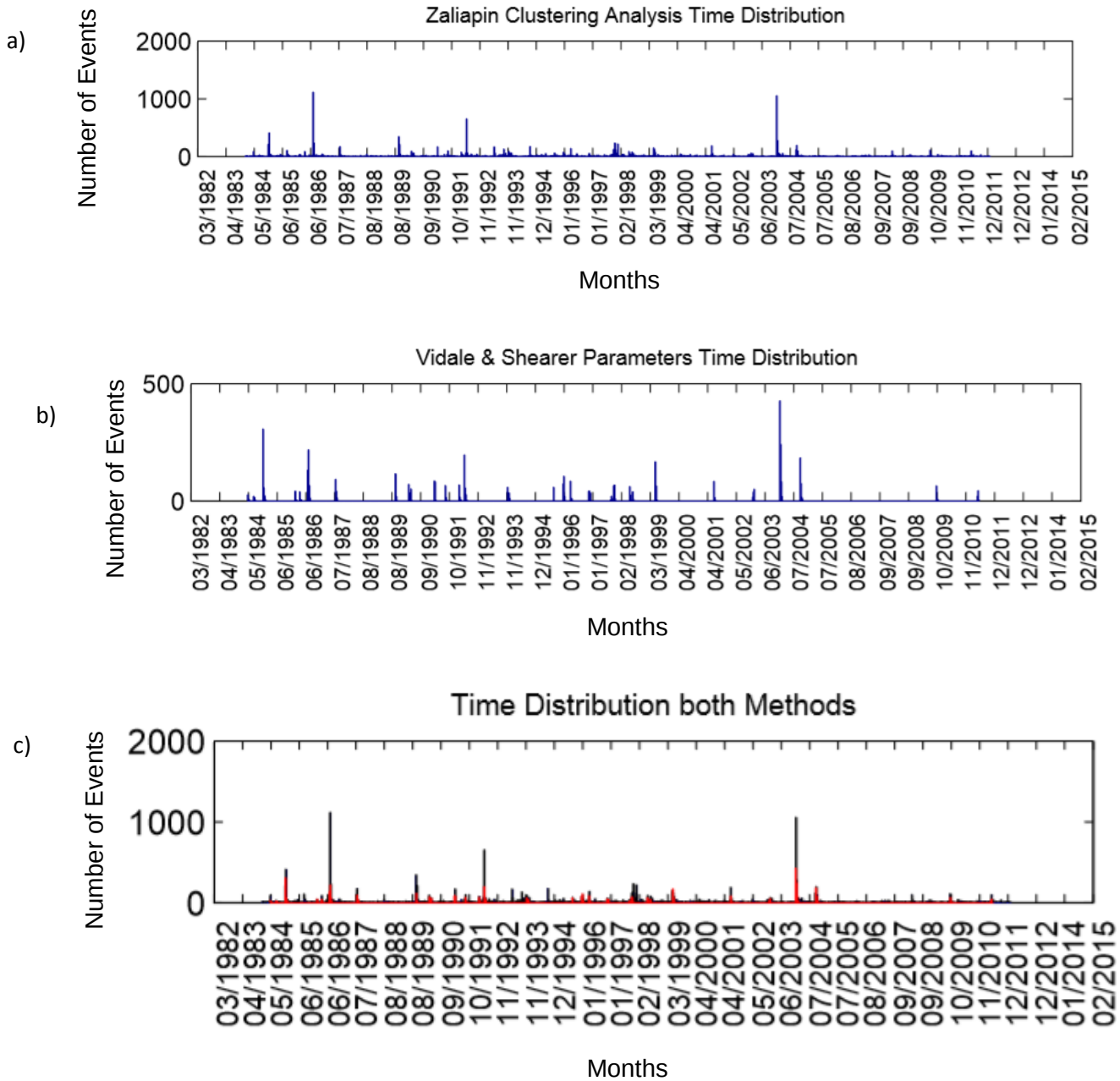
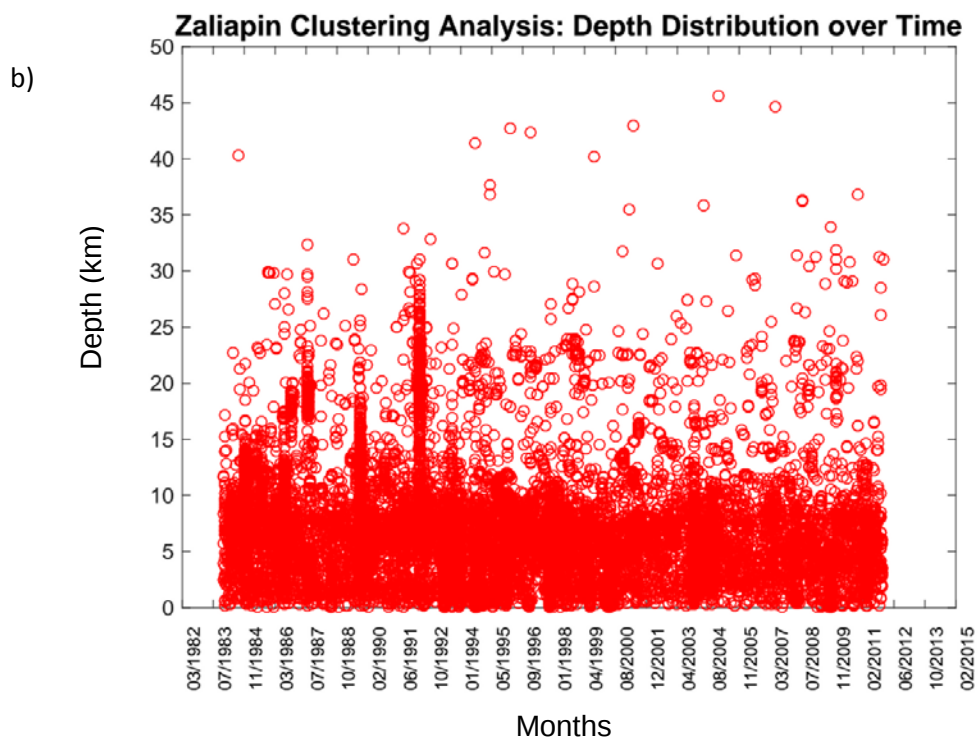
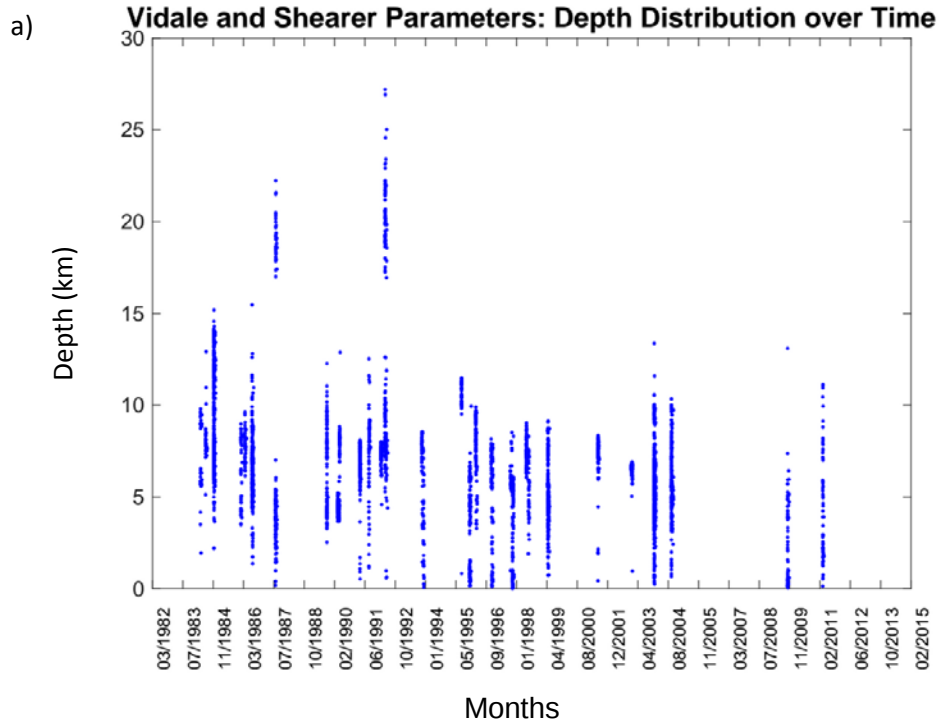


Fig 11. a) Event timing distribution for the Zaliapin method. b) Event timing distribution for the V&S method. c) Event timing for both methods, Zaliapin in blue and V&S in red. Note that the peaks, which correspond to swarm sequences, in the V&S distribution do not reach that of the Zaliapin distribution.



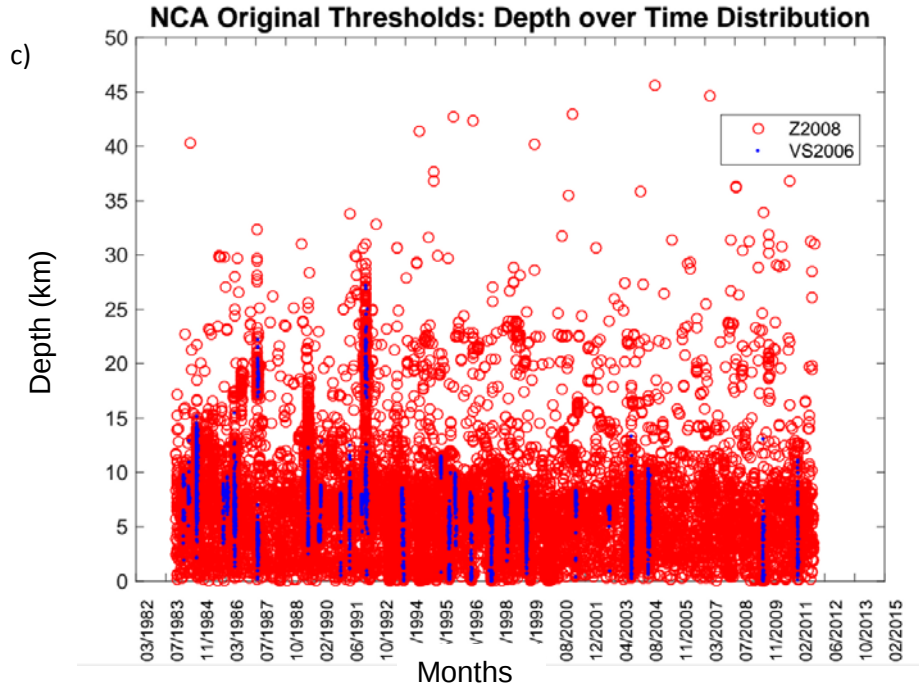
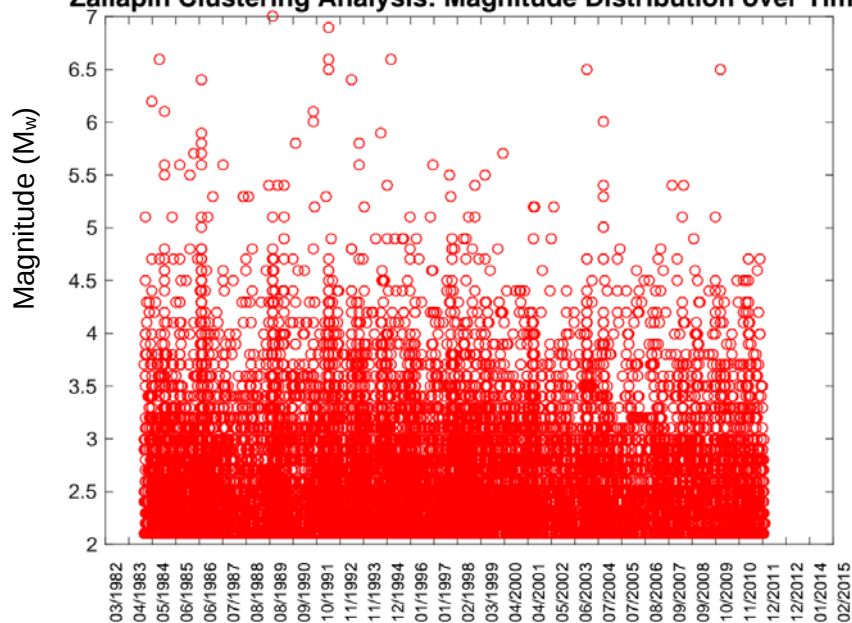
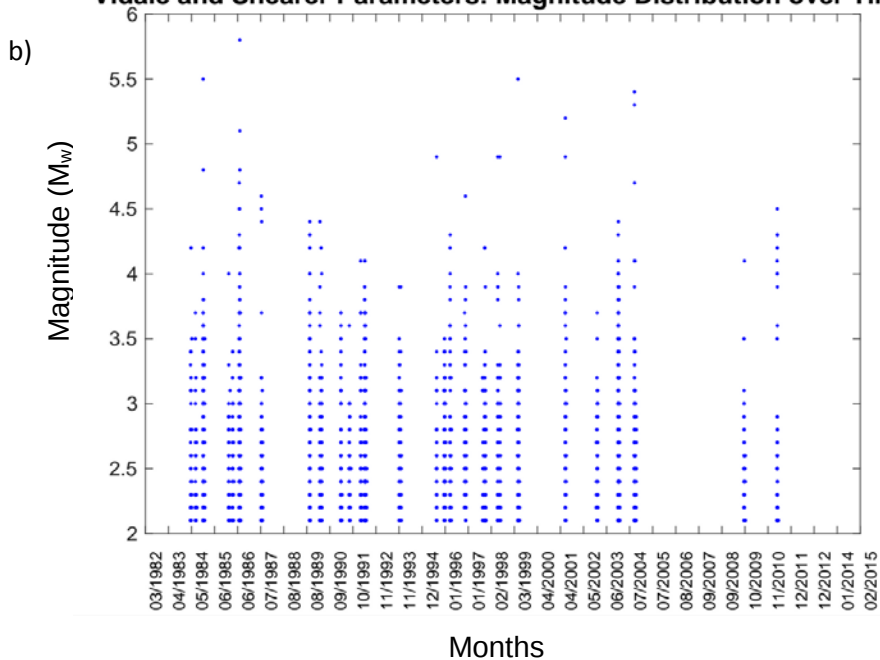


Fig. 12. a) Depth over time distribution for the Zaliapin method analysis. b) Depth over time distribution for the V&S parameters analysis. c) Depth over time distribution for both methods, Zaliapin in red and V&S in blue. Note that the V&S parameters identify the same swarms, but not the same number of events as the Zaliapin method.

a) **Zaliapin Clustering Analysis: Magnitude Distribution over Time**



b) **Vidale and Shearer Parameters: Magnitude Distribution over Time**



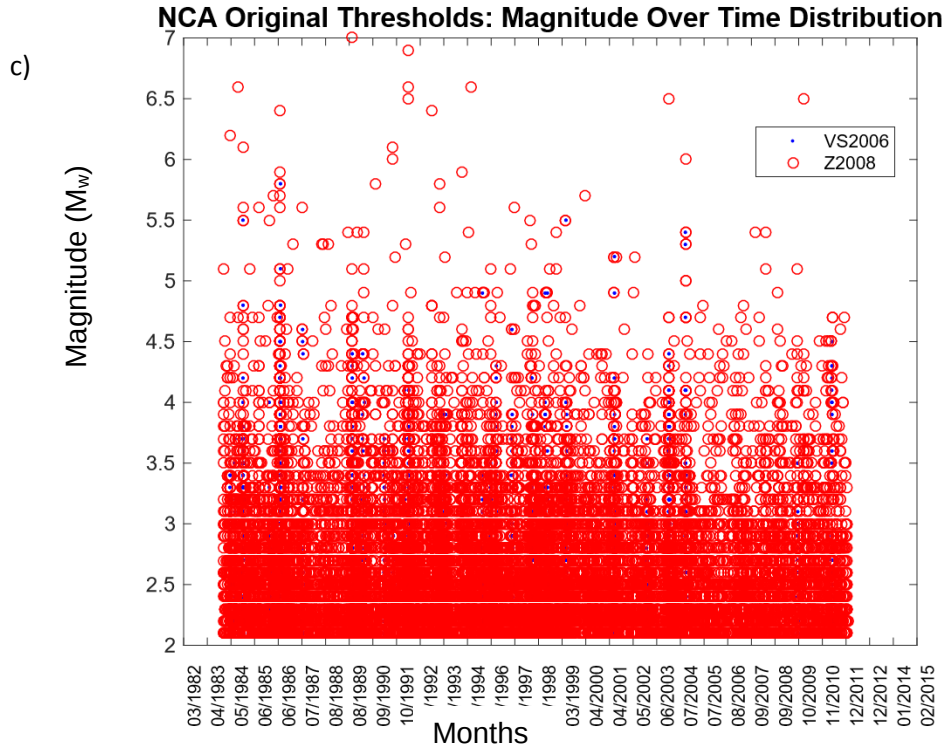


Fig. 13. a) Magnitude over time distribution for the Zaliapin method analysis. b) Magnitude over time distribution for the V&S parameter analysis. c) Magnitude over time distribution for both methods, Zaliapin in red and V&S in blue. Note that although V&S identifies the same swarm sequences, it doesn't identify the same number of events constituent of each swarm that Zaliapin does.

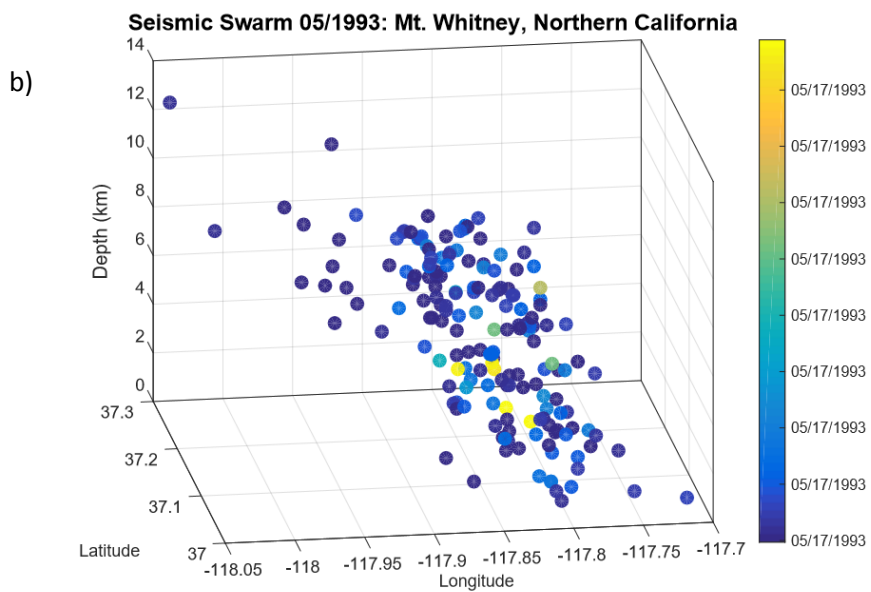
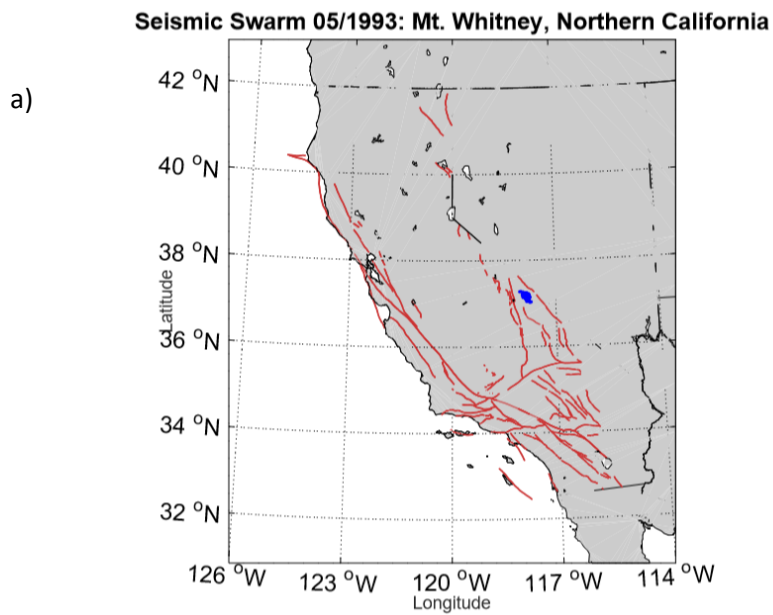
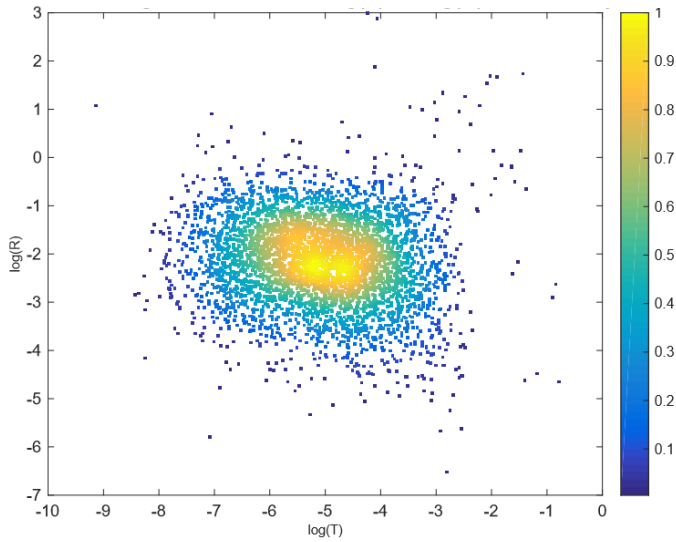


Fig. 14. a) Map of Northern California with fault lines and state borders plotted. Points in blue are events constituent of the 1993 seismic swarm near Mt. Whitney in Northern California, which was identified by Z2008, but not by VS2006 due to spatial constraints in the thresholds. b) Seismic swarm plotted over longitude, latitude and depth. Note that most events occur towards the beginning of the swarm, consistent with magma migration, with a few events happening towards the end.

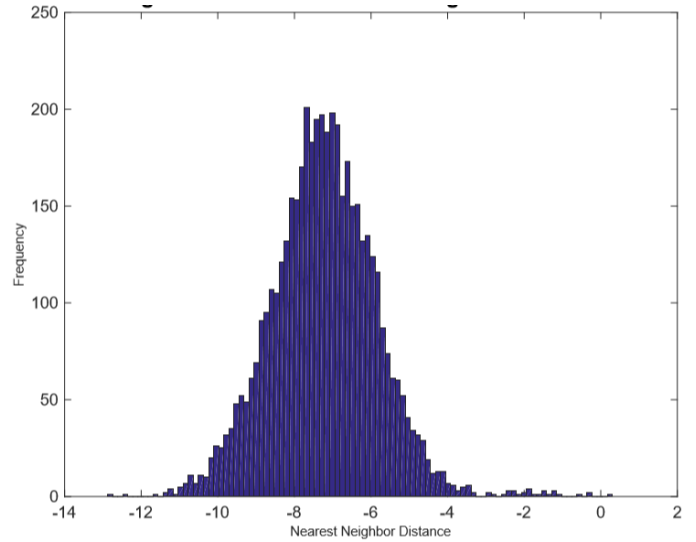
NCA Original Thresholds: $\log(T)$ vs $\log(R)$

a)



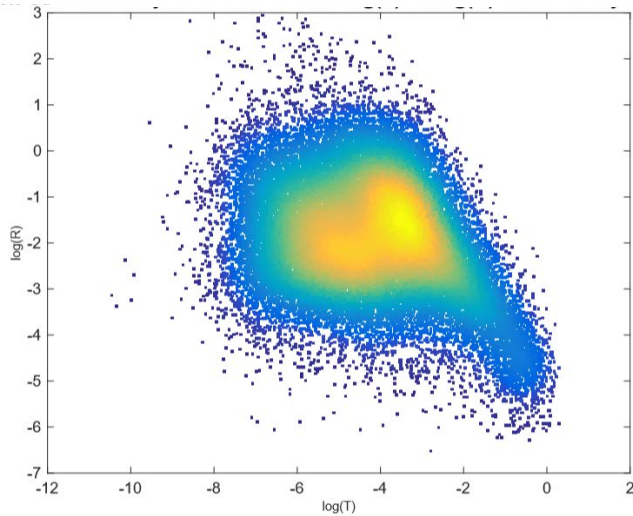
NCA Original Thresholds: Nearest Neighbor Distribution

b)



NCA Adjusted Thresholds: $\log(T)$ vs $\log(R)$

c)



NCA Adjusted Thresholds: Nearest Neighbor Distribution

d)

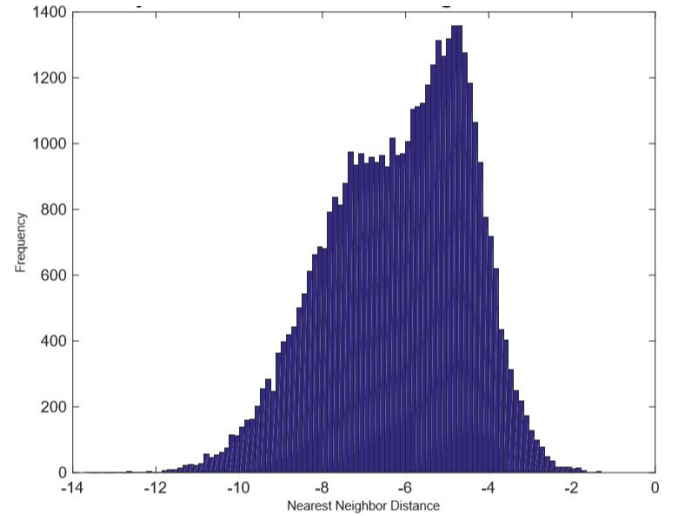


Fig. 15. a) Log of the interoccurrence time plotted against distance for the original thresholds in Northern California. b) Nearest neighbor distance distribution for the original thresholds in Northern California. c) Log of the interoccurrence time plotted against distance for the adjusted thresholds in Northern California. d) Nearest neighbor distance distribution for the adjusted thresholds in Northern California. Note that although the original thresholds identify far fewer events in the non-clustered distribution as well as less linearly related events in the $\log(T)$ vs $\log(R)$ plot, the adjusted thresholds identify a far higher number of events being clustered and potentially constituent of swarms.

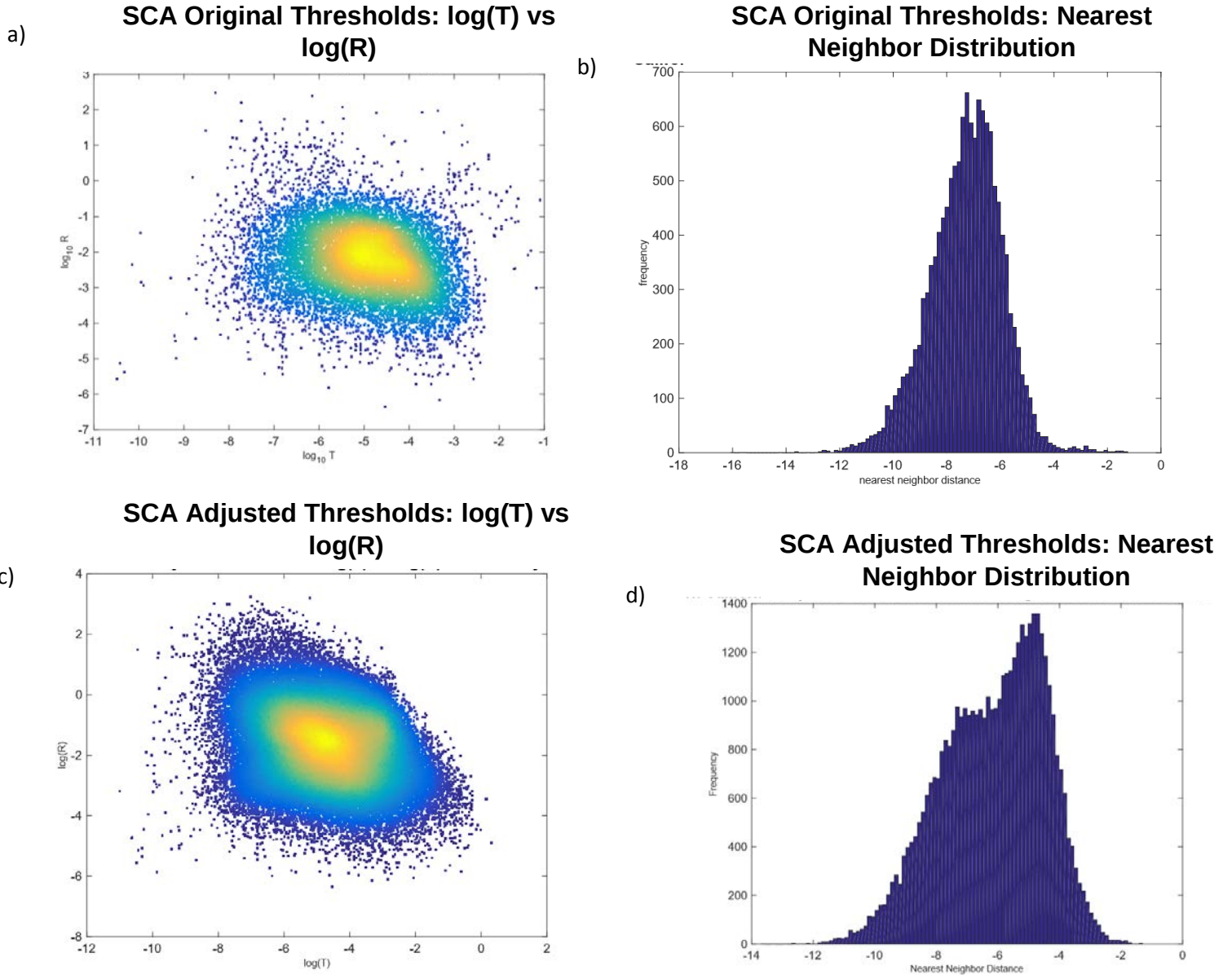


Fig. 16. a) Log of the interoccurrence time plotted against distance for the original thresholds in Southern California. b) Nearest neighbor distance distribution for the original thresholds in Southern California. c) Log of the interoccurrence time plotted against distance for the adjusted thresholds in Southern California. d) Nearest neighbor distance distribution for the adjusted thresholds in Southern California. Note that although the original thresholds identify far fewer events in the non-clustered distribution as well as less linearly related events in the $\log(T)$ - $\log(R)$ plot, the adjusted thresholds identify a far higher number of events being clustered and potentially constituent of swarms, consistent with the results from Northern California. In addition, the $\log(T)$ - $\log(R)$ plot of the adjusted thresholds demonstrates less of a bimodal distribution than exhibited in Northern California.

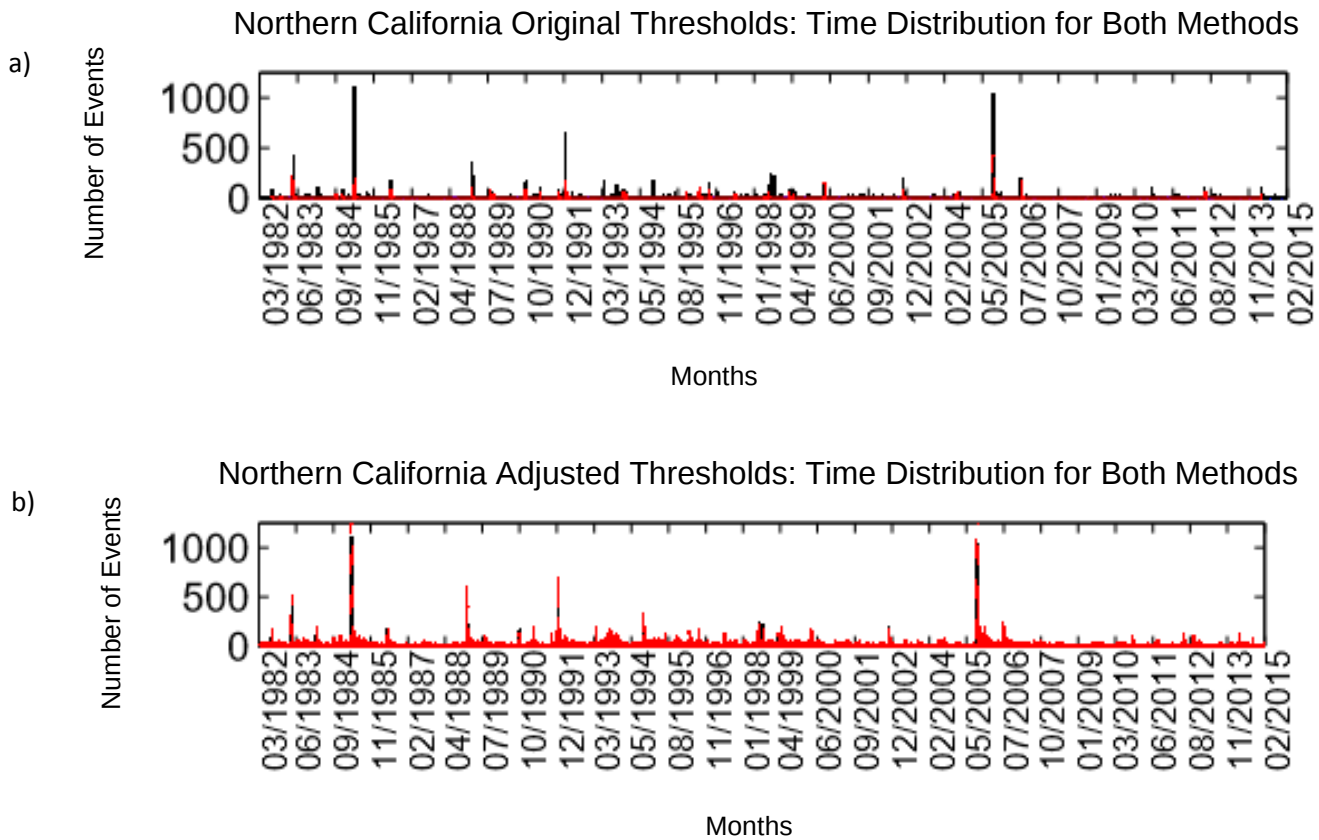
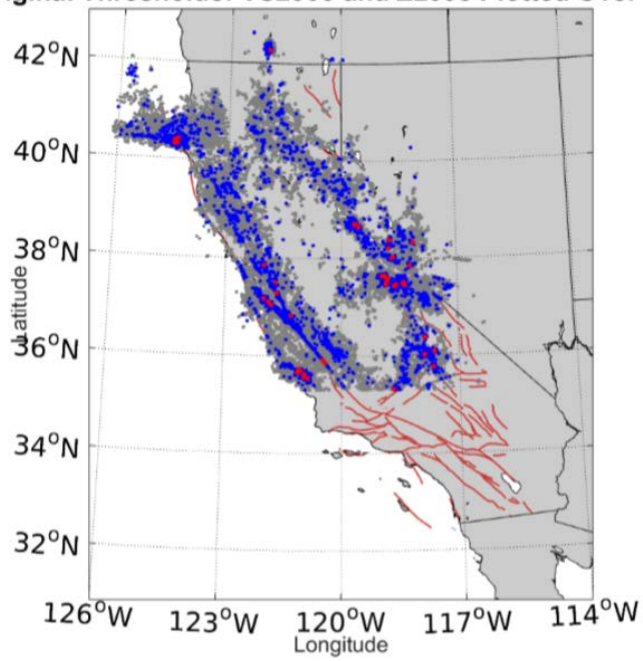


Fig. 17. a) Time distribution of the number of events spread over months between 1984-2011. Plotted in blue are the events and clusters identified by Z2008, in red are the events and clusters identified by VS2006 original thresholds. b) Time distribution of the number of events between 1984-2011. Plotted in blue are the events and clusters identified by Z2008, in red are the clusters and events identified by the adjusted thresholds. Note that not only are more clusters identified in b) than a, but more events overall are identified by the adjusted thresholds than the clustering analysis.

NCA Original Thresholds: VS2006 and Z2008 Plotted Over All Events

a)



NCA Adjusted Thresholds: VS2006 and Z2008 Plotted Over all Events

b)

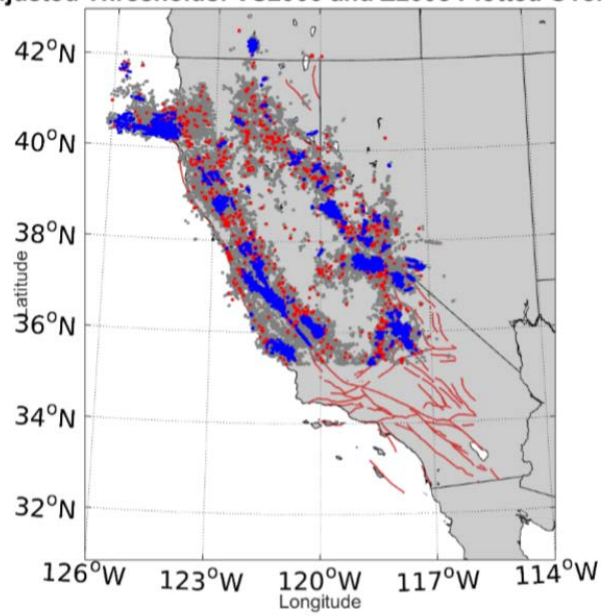
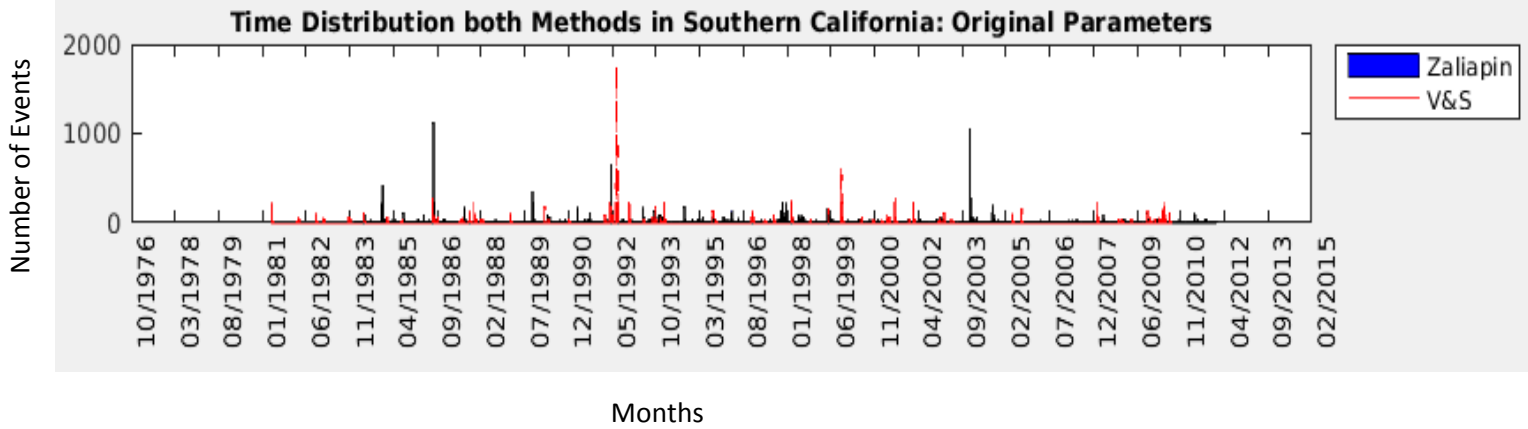


Fig. 18. a) Clusters of events plotted over Northern California for the original thresholds. In grey are all the earthquake events contained in Waldhauser's (2013) catalog, in red are all the events identified by Z2008, and in blue are the events identified by the original VS2006 thresholds. b) Clusters of events plotted over Northern California for the adjusted thresholds. In grey are all the events, in red are the events identified by Z2008, and in blue are the events identified by the adjusted VS2006 thresholds. Note that as the thresholds are expanded, the events expand from isolated bursts of seismic activity to longer trends of activity along major fault lines. In addition, more events are identified by the adjusted thresholds, expanding beyond the number of events identified by the clustering analysis.

a)



b)

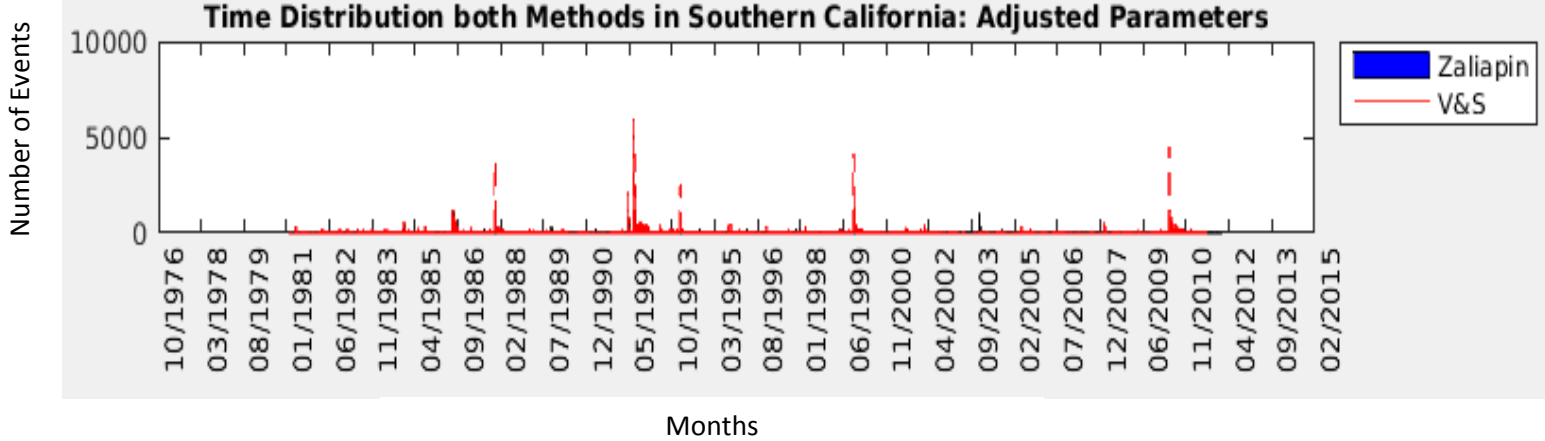


Fig. 19. a) Time distribution for the number of swarm events occurring in Southern California over months between 1984-2011. Plotted in blue are the clusters and events identified by Z2008, plotted in red are the clusters and events identified by the original thresholds of VS20006. b) Time distribution for the number of events occurring over months between 1984-2011. Plotted in blue are the clusters and events identified by Z2008, plotted in red are the clusters and events identified by the adjusted thresholds. Note that not only do the adjusted thresholds identify more clusters than the original thresholds, but identify more events overall than does the clustering analysis.

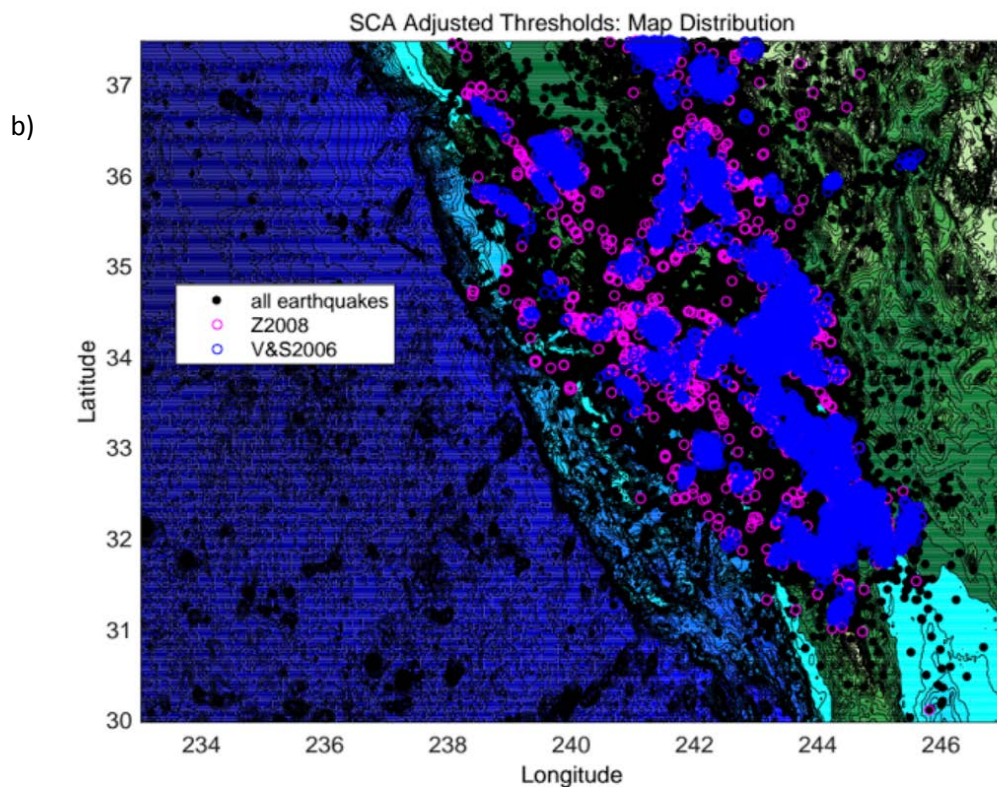
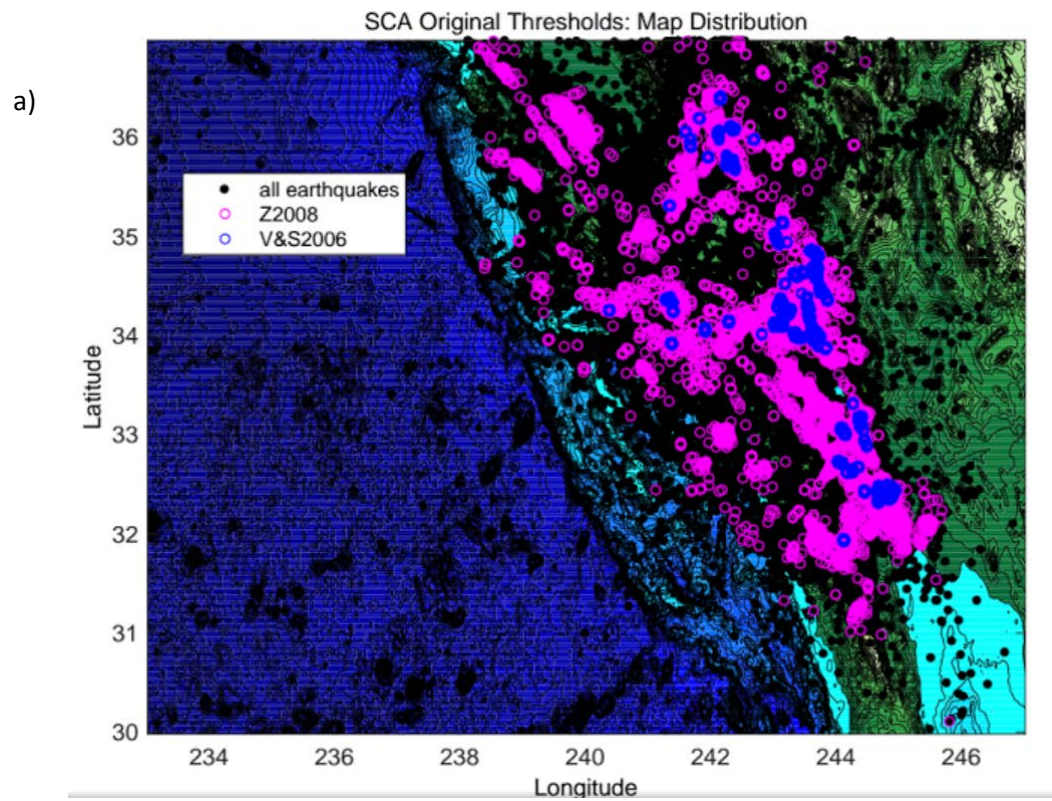
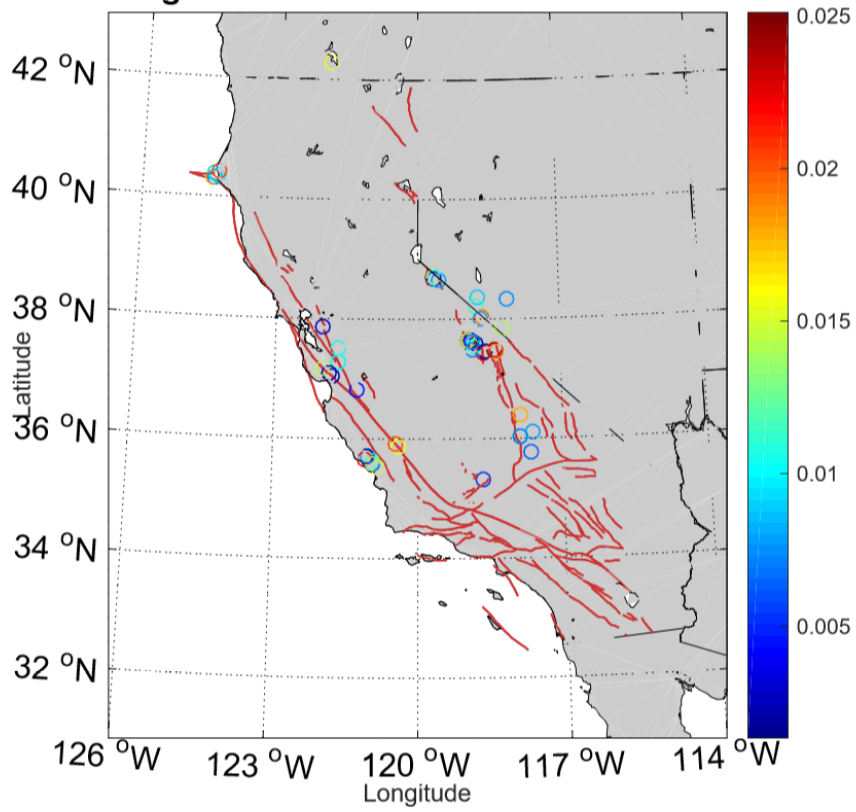


Fig. 20. a) Events identified by the original VS2006 thresholds plotted over Southern California. In black are all the events contained in Hakusson, Yang and Shearer's (2013) catalog, in purple are the events identified by Z2008, and in blue are the events identified by VS2006. b) Events identified by the adjusted thresholds plotted in Southern California. In black are all the events, in purple are the events identified by Z2008, and in blue are the events identified by the adjusted thresholds. Note that the clusters adjust from isolated bursts of seismicity to longer, linear trends of seismicity along major fault zones in Southern California, as well as contain more events in a greater number of locations between the original and adjusted thresholds.

NCA Original Thresholds: Mean Swarm Locations

a)



NCA Adjusted Thresholds: Mean Swarm Locations

b)

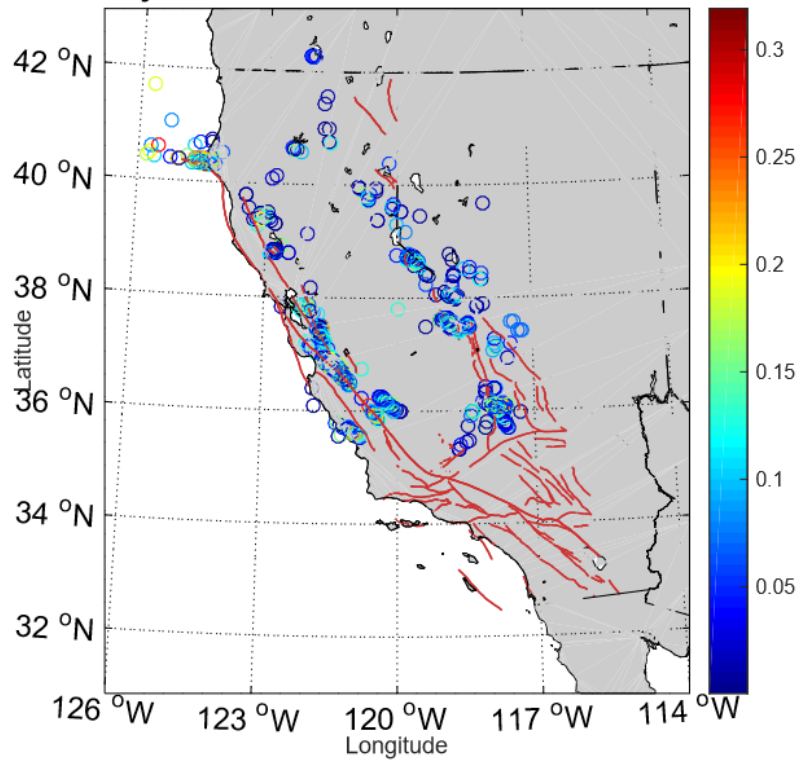


Fig. 21. a) Mean locations of clusters of events identified by the original thresholds in Northern California. Color bar corresponds to the distance traveled between the first and last event in each swarm. b) Mean locations of the events identified by the adjusted thresholds in Northern California, color bar corresponds to the distance traveled between the first and last event in each swarm. Note that there are more swarms identified by the adjusted thresholds than the original thresholds, and a larger fraction of these swarms appear to migrate. In addition, many swarms in both maps appear to be related, indicating that the thresholds could be separating them into individual clusters, rather than locating them all as one large cluster.

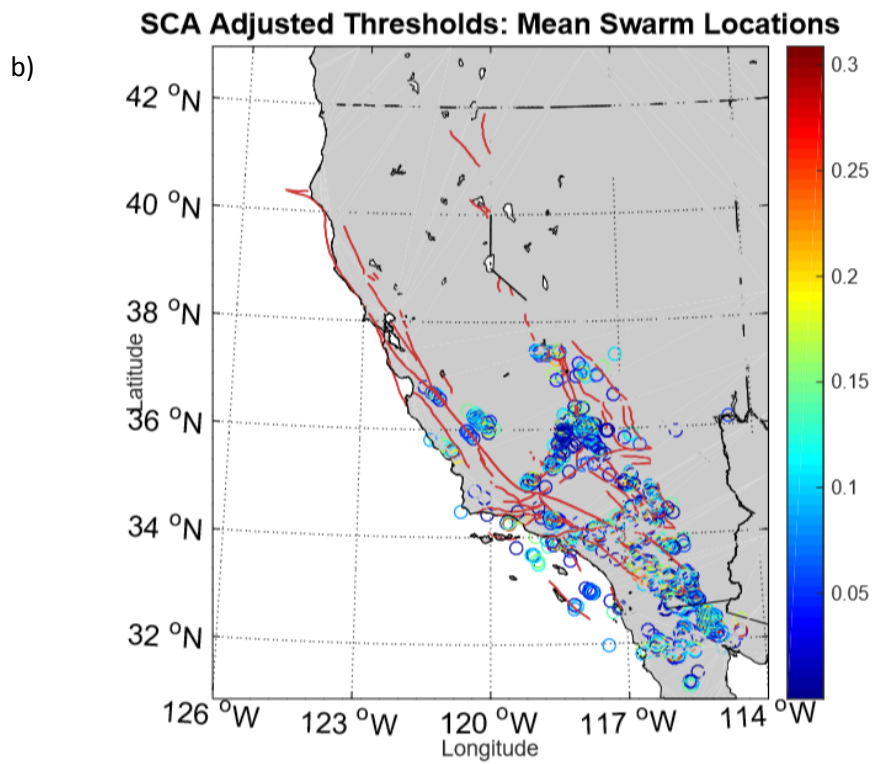
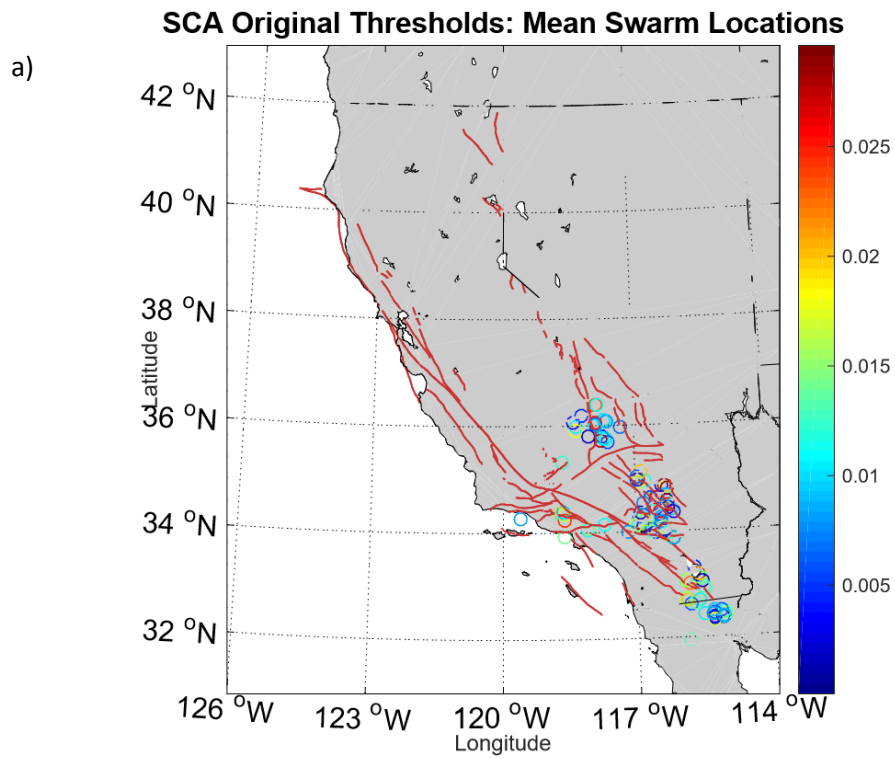
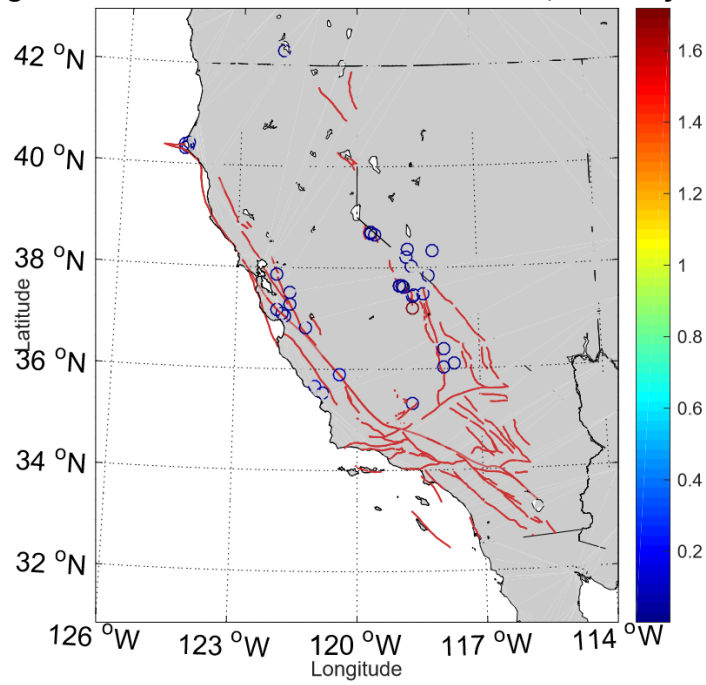


Fig. 22. a) Mean locations of clusters of events identified by the original thresholds in Southern California. Color bar denotes distance between starting and end location of each cluster b) Mean locations of the events identified by the adjusted thresholds in Southern California, color bar denotes distance between starting and end location of each swarm. Note that there are more swarms identified by the adjusted thresholds than the original thresholds, and a larger fraction of these swarms appear to migrate. In addition, many swarms in both maps appear to be related, indicating that the thresholds could be separating them into individual clusters, rather than locating them all as one large cluster. Finally, clusters in Southern California appear to be more constrained in terms of area than clusters in Northern California, which trend over the entire region.

NCA Original Thresholds: Mean Swarm Locations, Manually Scrubbed

a)



NCA Adjusted Thresholds: Mean Swarm Locations, Manually Scrubbed

b)

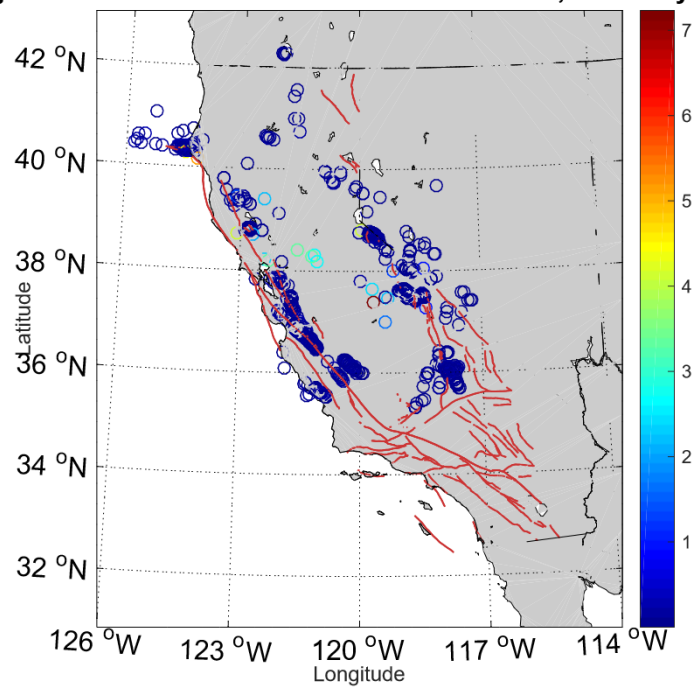
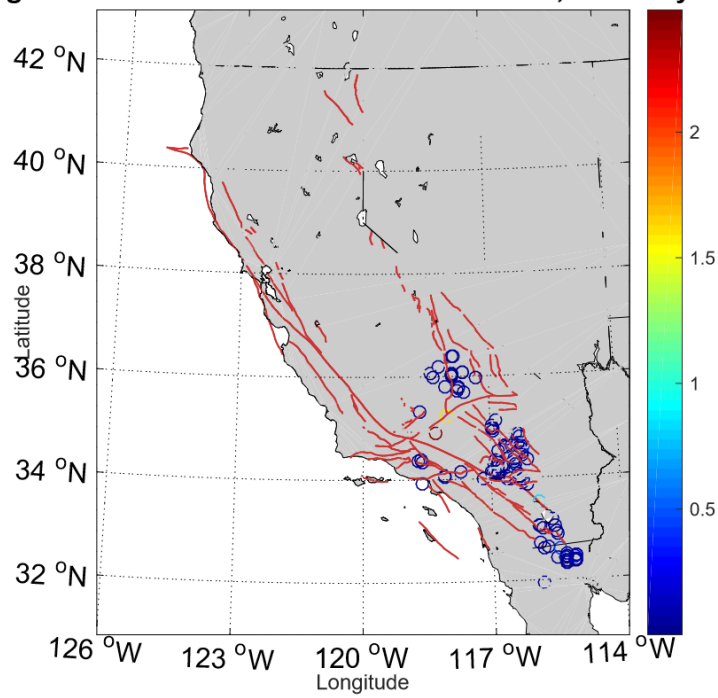


Fig. 23. a) Merged and scrubbed swarm locations for the original thresholds in Northern California. Color bar denotes distance traveled between first and last event in each swarm.
b) Merged and scrubbed swarm locations for the adjusted thresholds in Northern California. Note that even with the merging of clusters and aftershocks removed, the adjusted thresholds still identify more swarms and more events constituent of swarms.

a) **SCA Original Thresholds: Mean Swarm Locations, Manually Scrubbed**



b) **SCA Adjusted Thresholds: Mean Swarm Locations, Manually Scrubbed**

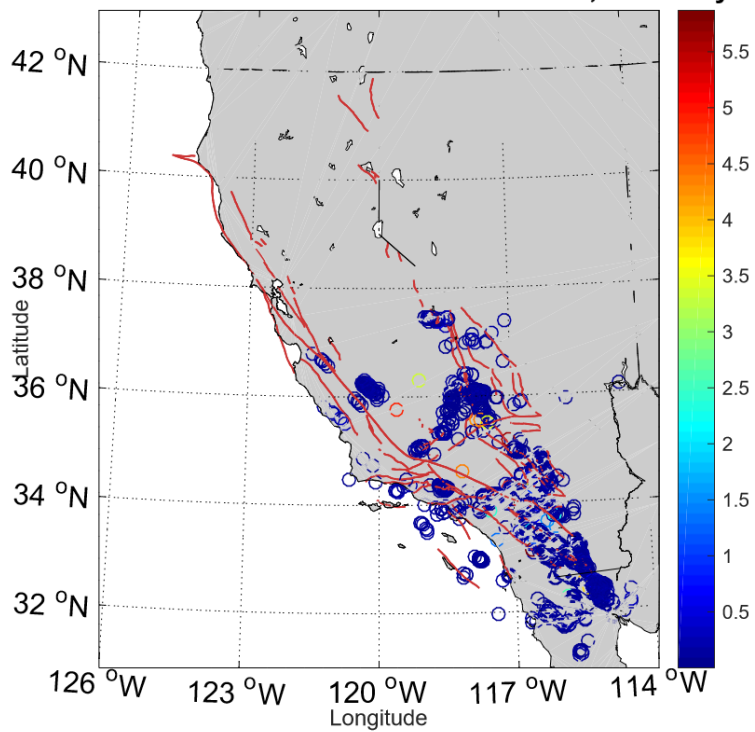
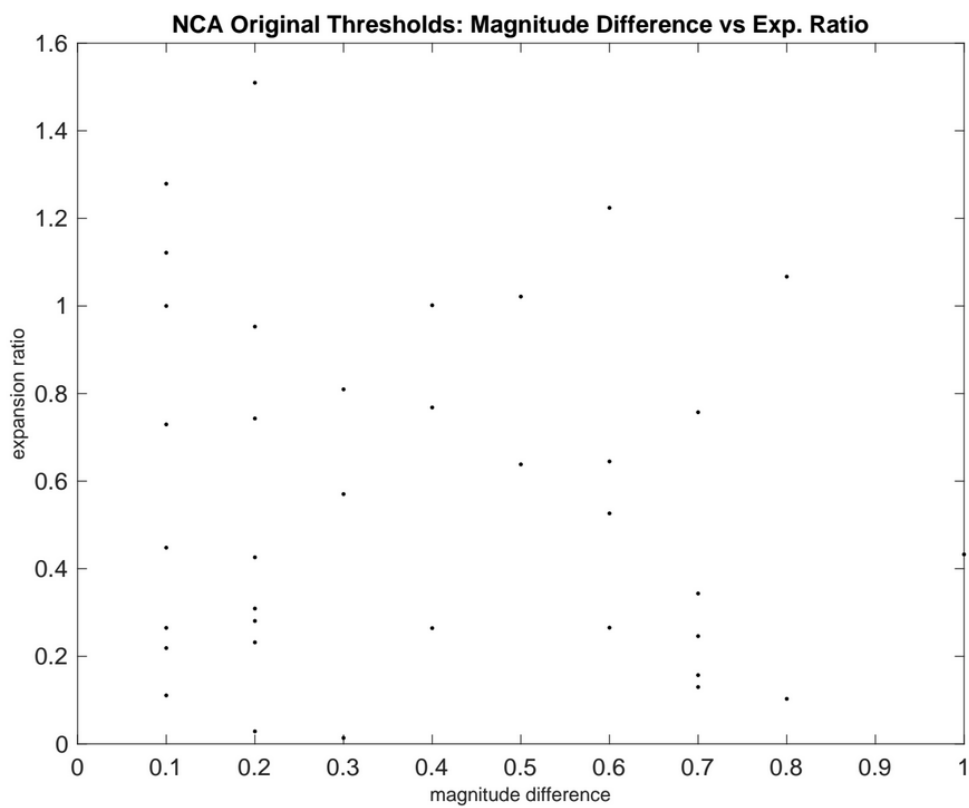


Fig. 24. a) Merged and scrubbed swarm locations for the original thresholds in Southern California. Color bar denotes distance traveled between first and last event in each swarm.
b) Merged and scrubbed swarm locations for the adjusted thresholds in Southern California. Note that even with the merging of clusters and aftershocks removed, the adjusted thresholds still identify more swarms and more events constituent of swarms.

a)



b)

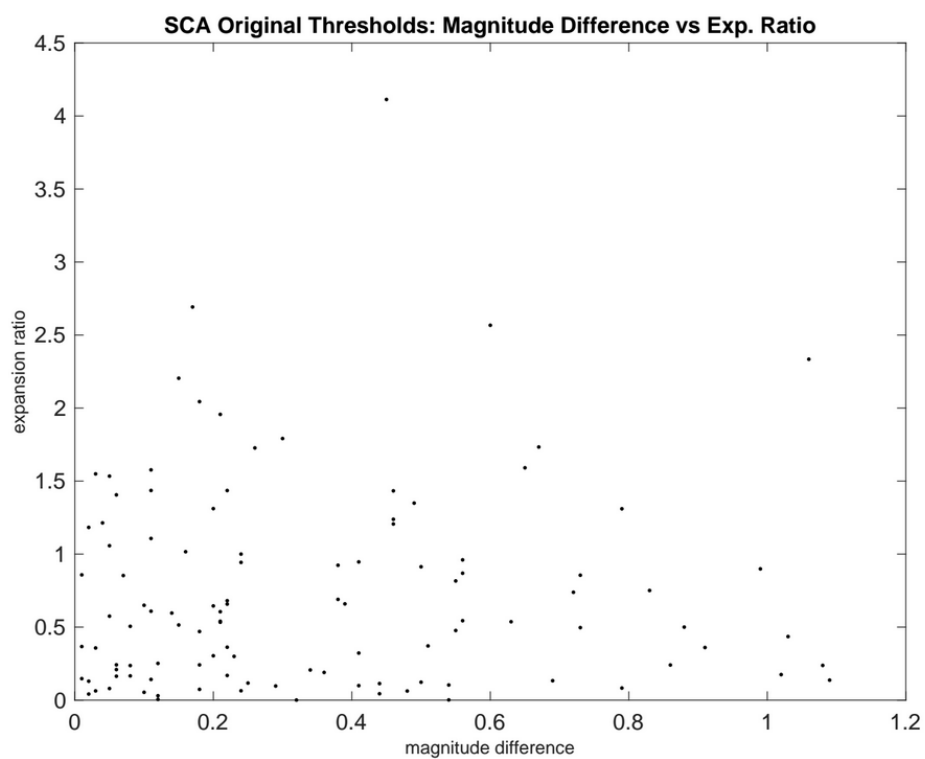


Fig. 25. a) Expansion ratio plotted against the magnitude difference for the original thresholds catalog in Northern California. b) Expansion ratio plotted against the magnitude difference for the original thresholds catalog in Southern California. Note that for both plots, there is no definite general trend in the distribution of the points, as well as most of the points being clustered towards the beginning, rather than evenly distributed throughout.

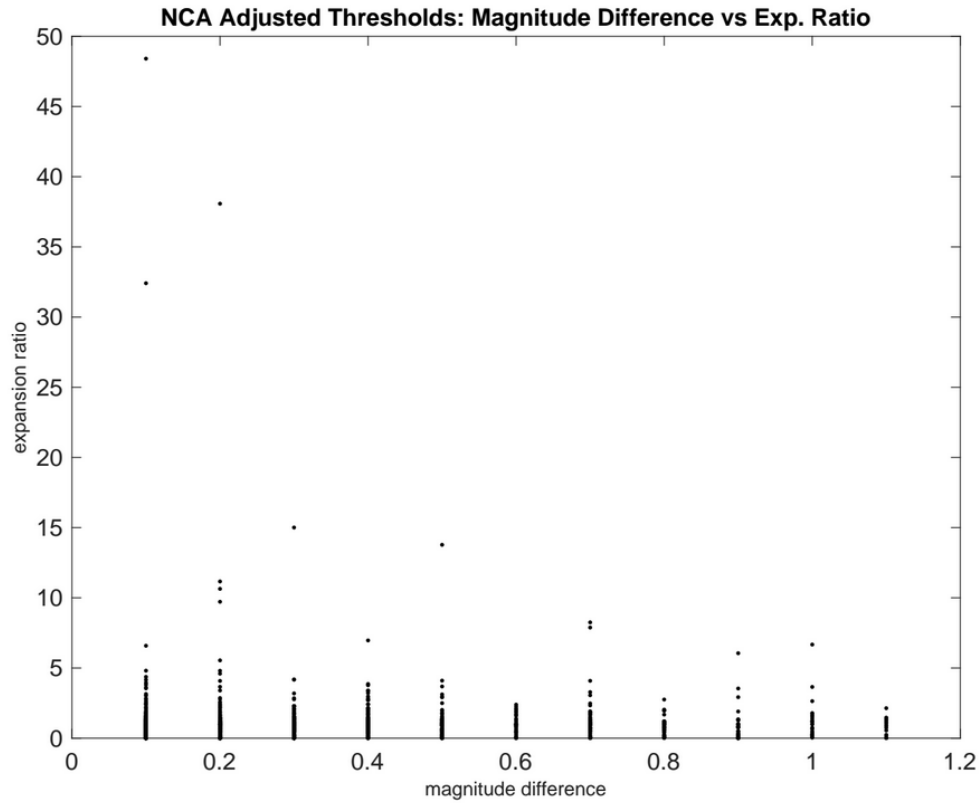


Fig. 26. Expansion ratio plotted against the magnitude difference for the adjusted thresholds catalog in Northern California. Note the definite exponential decay of increasing magnitude difference with decreasing expansion ratio, which differs from either of the original thresholds catalogs.