UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

A RETROSPECTIVE, MACHINE-LEARNING ASSISTED ANALYSIS OF SEISMIC SEQUENCE MIGRATIONS IN OKLAHOMA, AND DELINEATING FAULT STRUCTURES

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

KAYCEE SIMS

Norman, Oklahoma

2022

A RETROSPECTIVE, MACHINE-LEARNING ASSISTED ANALYSIS OF SEISMIC SEQUENCE MIGRATIONS IN OKLAHOMA, AND DELINEATING FAULT STRUCTURES

A THESIS APPROVED FOR

SCHOOL OF GEOSCIENCES

BY THE COMMITTEE CONSISTING OF

Dr. Jacob Walter, Co-chair

Dr. Brett M. Carpenter, Co-chair

Dr. Xiaowei Chen

ACKNOWLEDGEMENTS

Thank you to the committee for this thesis for the time and effort involved in accomplishing this project.

Thank you to Dr. Jake Walter for answering my extensive list of questions over the last two years, as well as the advice and encouragement.

Thank you, Mom and Dad, for everything you've done to support my dreams. A line in a paper is nowhere near enough to reflect it all.

Thank you, Kyle, for putting up with me all this time and keeping me (mostly) sane.

Thank you, Renae, for support and feedback along the way.

Thank you to the Oklahoma Geological Survey and the USGS for providing data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Seismogenic fault geometries are important to understanding the seismic potential of fault slip in natural systems. Their orientation relative to the principal tectonic stress directions control their susceptibility to slip. While fault orientations can be ascertained through detailed field mapping or subsurface imaging, rapid automatic identification of fault structures may be useful for future seismic hazard analysis. We apply a methodology workflow that can be automated and utilize central Oklahoma seismicity since 2010 as a test of the concept. To conduct this retrospective analysis, we utilize a Python package, easyQuake, that was recently developed by the Oklahoma Geological Survey (OGS), to augment their earthquake monitoring via pickers that are trained on millions of seismograms with neural network deep learning. The easyQuake software was used to analyze seismic data collected from 2009 to 2015 - specifically to investigate the possible migration of earthquake nucleation in the Jones, Oklahoma region - as well as increase the catalog completeness and accuracy. Jones is the area of interest for this study as it was an early case study for large volume (~1 Mbbl/mo) wastewater injection inducing earthquakes at greater than 10 km distances. The software outputs a catalog in QuakeML format, of which the associated metadata, such as pick times, can be read into hypoDD format. The catalog is then relocated with hypoDD to explore cluster and fault geometries, as well as earthquake sequence migration. With the hypoDD results, we utilize machine-learning techniques packaged in the Python machine-learning library scikit-learn for spatial clustering (e.g., DBSCAN) followed by linear regression (e.g., RANSAC) of those clusters to identify suspected fault segments. We identified potential instances of previously unmapped, seismically active segments in the Jones area and plan to expand the analysis to

examine a wider range of the state's fault segments and more current seismic sequences. Many of the segments mapped would not have been identifiable without the expansion of the microseismic catalog easyQuake provides. This type of work improves our understanding of the seismic hazard as it increases the database of known fault segments throughout the broader region and shows that these seismogenic fault segments can be identified automatically.

PLAIN LANGUAGE SUMMARY

We study how central Oklahoma's fault systems may be organized. We begin with data from 2010, during a time of higher-than-normal seismic activity. To study these events and subsurface features today, using historical data, we use easyQuake, a Python package recently developed by the Oklahoma Geological Survey. This package uses machine learning techniques to supplement the earthquake monitoring systems already in place. Data from 2009 to 2015 was analyzed using this new software to make the catalogs both more complete and more accurate, and to look at how earthquakes developed in the Jones, Oklahoma region. The software creates a catalog file in QuakeML format; this allows the associated metadata, such as pick times, to be read into the hypoDD format. Jones was an early case study for large volume wastewater injection which was potentially causing earthquakes. HypoDD is used to relocate the events, which allows us to explore cluster and fault geometries, as well as earthquake sequence migration with machine-learning techniques. Events are sorted into clusters, followed by linear regression of those clusters to identify suspected fault segments. This type of work improves our understanding of the seismic hazard as it increases the database of known fault segments throughout the broader region.

CHAPTER 1: INTRODUCTION

1.1 Introduction

What is causing increased seismicity in Oklahoma over the last decade has been a topic of interest (e.g., McNamara et al., 2015; Schoenball & Ellsworth, 2017). The level of connection with wastewater injection wells is not entirely known, though there has been a strong correlation between wastewater injection volumes and seismic activity in the region (e.g., Keranen et al., 2014; Walter et al., 2019). One of the avenues for induced seismicity explored in this study is the possible relationship between hydraulic loading levels and seismic activity as is studied in other parts of the world (e.g., Costain & Bollinger, 2010, Foulger et al., 2017, Grasso et al., 2018). Due to data limitations, among other reasons, it is difficult to prove definitively if these events were induced or naturally occurring. "Definitively" is always difficult, though, as these are subsurface features whose environment is difficult to replicate for laboratory experiments.

To increase the completeness and accuracy of available data, easyQuake is employed to create a catalog that can be further analyzed. easyQuake is an open-source Python package designed for automated data collection via pickers trained on millions of seismograms with neural network deep learning (Walter et al., 2020). The primary area of this study is central Oklahoma, focusing near Jones, OK as it is home to a massive wastewater injection site, pumping roughly 1Mbb/month (Keranen et al., 2014). Later in the study, the same techniques used for Jones area applied to a broader region.

Small earthquakes – below M2.5 – have often not been reported in official catalog databases as it has not been mandatory. Smaller events could be reported, but the USGS mandatory reporting cutoff for the lower 48 states was M2.5. A more complete microseismic catalog will examine smaller events and help fill in seismic hazard information for areas that are currently not covered. Looking at even just 2010 we see there are many events below M2.5 that went unreported by the USGS.

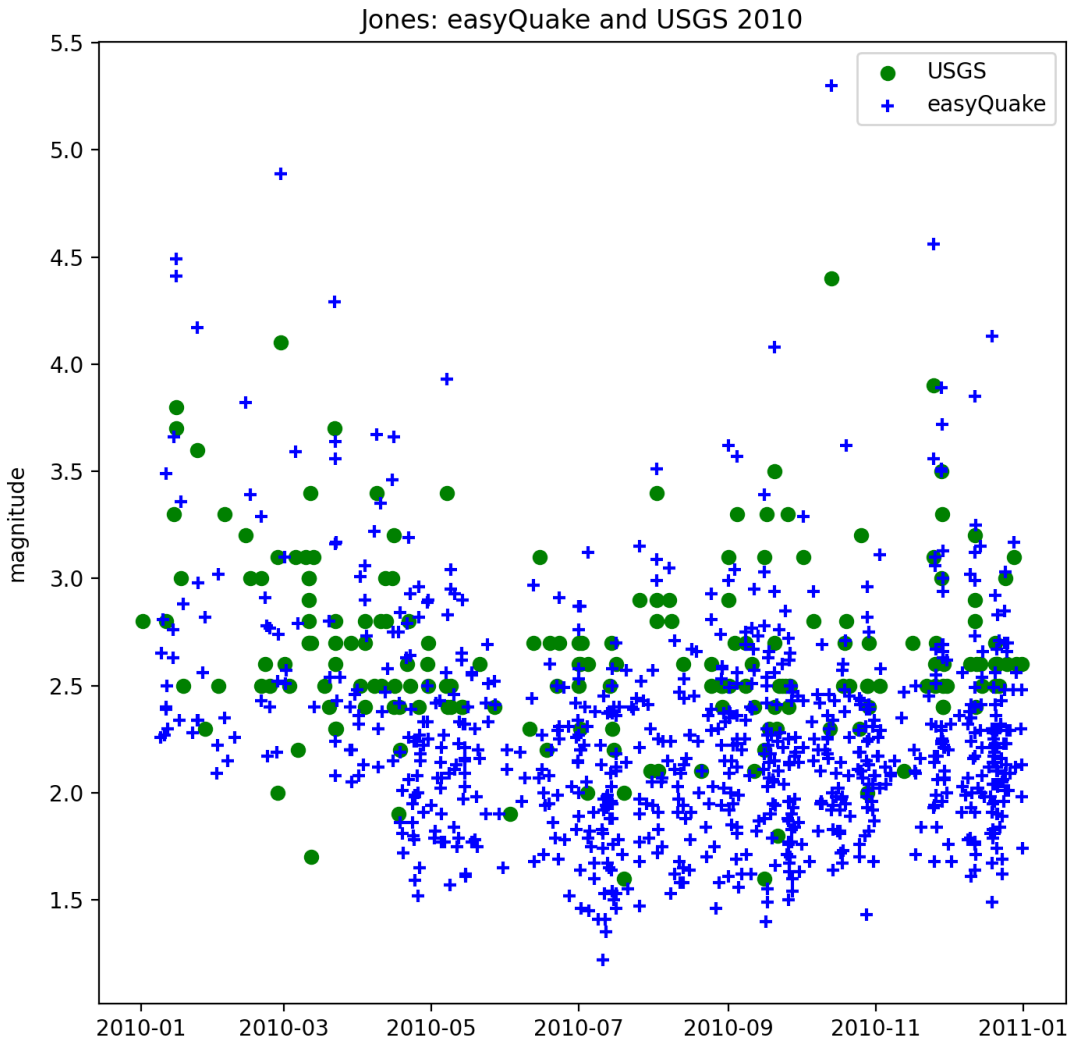


Fig 1.1.1. USGS events shown as green dots; easyQuake events shown as blue crosses. We see here that there are many more microseismic events occurring than are captured in the official catalog due to the M2.5 mandatory reporting cutoff.

Following a system based on and similar to the FaultID algorithm of Skoumal et al., 2019, preliminary fault maps are created by relocating and clustering seismic events, then using cluster orientations to delineate potential fault segments. Based on the origin times of events and their locations along proposed fault lines, we examine earthquake migrations and the relationship with pressure propagation from wastewater injection, similar to Keranen 2014.

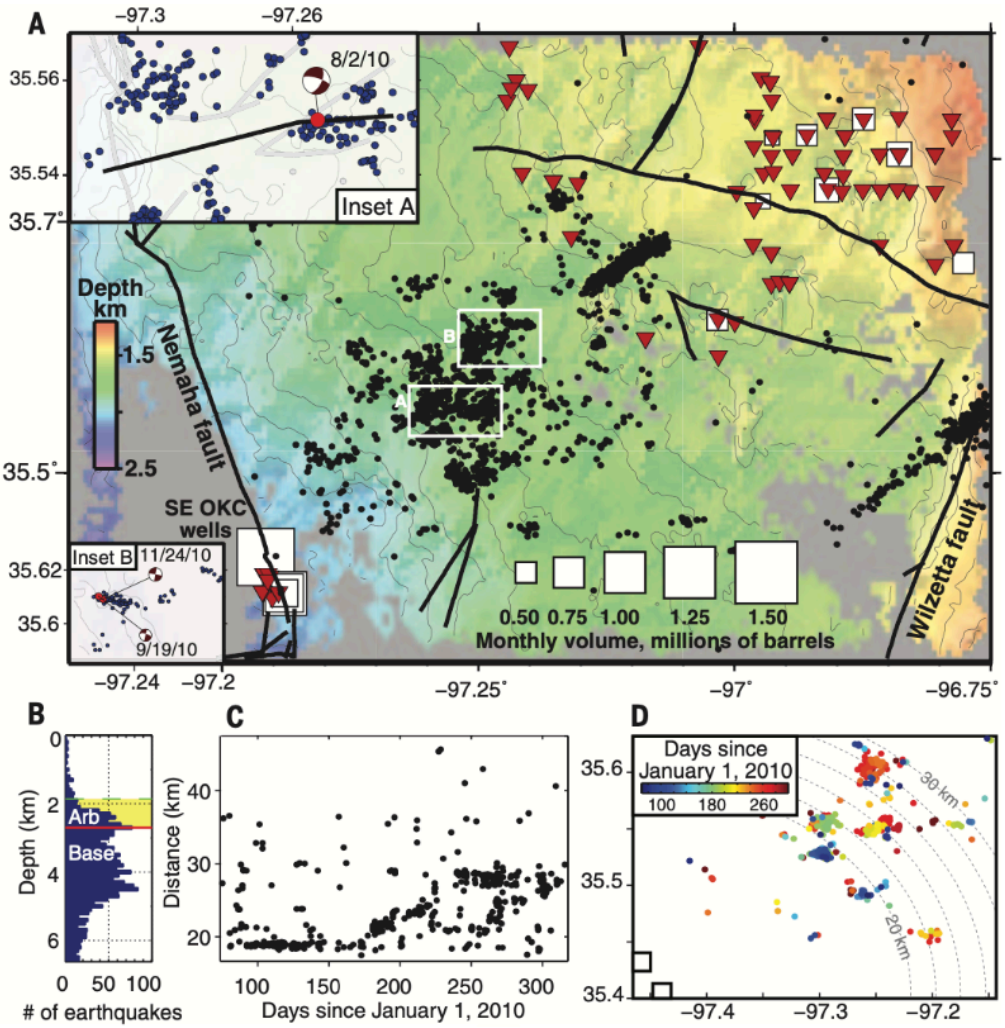


Fig 1.1.2. Earthquake catalog and swarm migration data. From Keranen et al., 2014. (a) Catalog of events with well locations, injection volumes, and known faults. (b) Event depths. (c) Distance from events March 2010-October 2010 to the Southeast OKC disposal wells. (d) March 2010-October 2010 event locations. This figure serves as a baseline for several aspects of this study. Primarily: the location of the high-volume wastewater injection wells, and the event-to-well distances.

1.2 Regional Background

Central Oklahoma is home to part of an active seismic zone that extends through the Texas panhandle; the earthquakes in this region are typically shallow compared to other seismically active regions (Walter et al., 2018). One of Oklahoma's major fault systems is the Nemaha fault system, which runs North-South and is surrounded by smaller faults (Lawson and Luza, 1995). Another major system is the Meers fault, which could produce an M7.0 or larger earthquake if the entire length were to rupture (Walter et al., 2019).

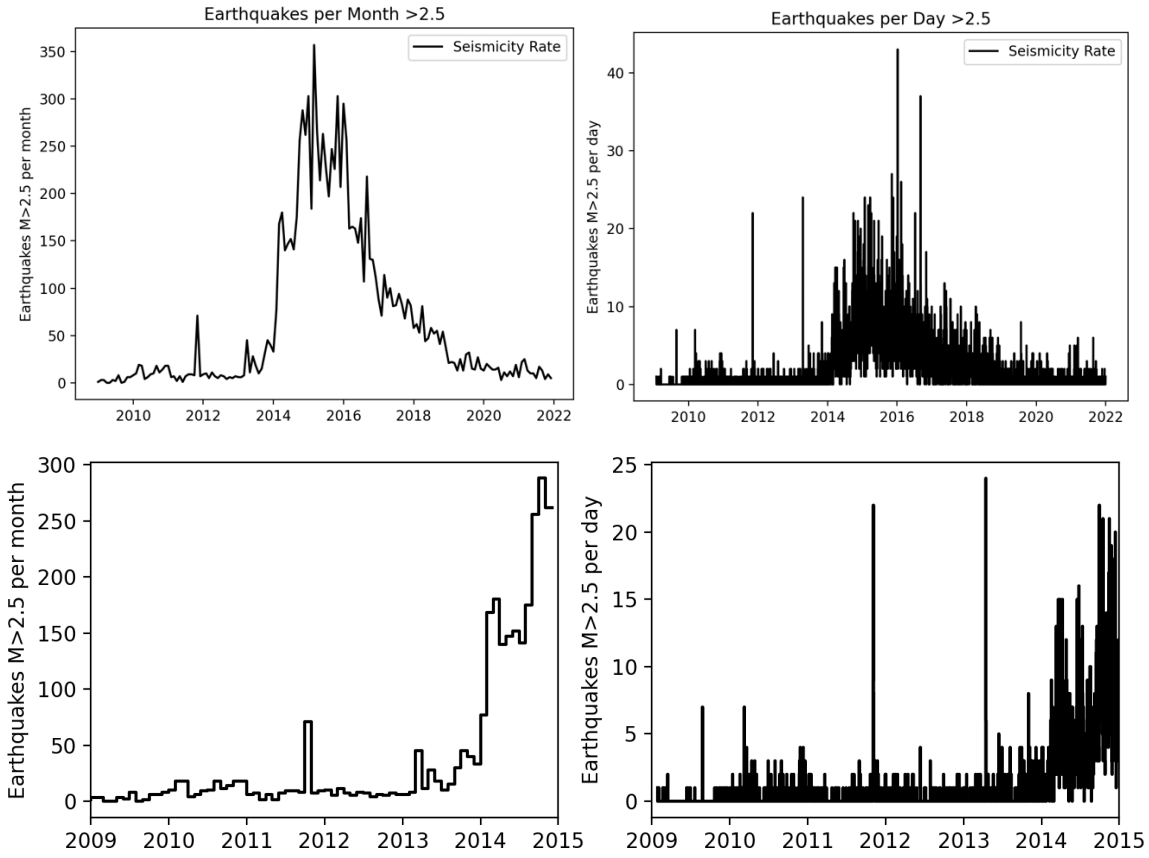


Figure 1.2.1 Comparison of seismic rates - monthly(left) and daily(right) - over time reported in the USGS catalog. Only events with a magnitude M2.5 or greater are reported. The top two panels show 2009 through 2021 data. The lower panels show 2009 through 2015 as this was the end of the period concurrent with the Keranen14 study. The past two years have seen a return of seismicity rates to levels near pre-2014 range.

The Jones area is just East of the Nemaha fault, and West of the Wilzetta fault, as shown in Figure 1.1.2. Prior to 2009, central Oklahoma was seismically quiet compared to the following years (e.g., Skoumal et al., 2019; Schoenball et al., 2017, Schoenball et al., 2018). Figure 1.2.1 shows the USGS reported seismic events for 2009 through 2015. As shown here there were generally fewer than five events M2.5 or above per day until 2013. Before 2009 there were even fewer. In 2015 though there were up to 20 events M2.5 or above, which is in stark contrast to the prior years. Four of the five 'large' earthquakes - M5.0 or greater - in this region have been in the last twelve years, as well as three events just under M5.0 which were either foreshocks or aftershocks of these events (Walter et al. 2019) as shown in Table 1.2.1.

Date (yyyy/mm/dd)	Name	Magnitude	County
2016/09/03	Pawnee earthquake	5.8	Pawnee
2011/11/06	Prague earthquake	5.7	Lincoln
1952/04/09	El Reno earthquake	5.5	Canadian
2016/02/13	Fairview earthquake	5.1	Woods
2016/11/07	Cushing earthquake	5.0	Payne
1882/10/22	Choctaw Nation earthquake	4.8	Southeast Oklahoma
2011/11/05	Prague foreshock	4.8	Lincoln
2016/01/07	Fairview foreshock	4.8	Woods
2011/11/08	Prague aftershock	4.8	Lincoln
2015/11/19	Alfalfa County earthquake	4.7	Alfalfa

Table 1.2.1 Significant Oklahoma Earthquakes, organized by Magnitude. From Walter et al., 2019. The region has seen more large events in the past decade than would be expected from the regional history. Of the ten events listed in this table, only two occurred prior to

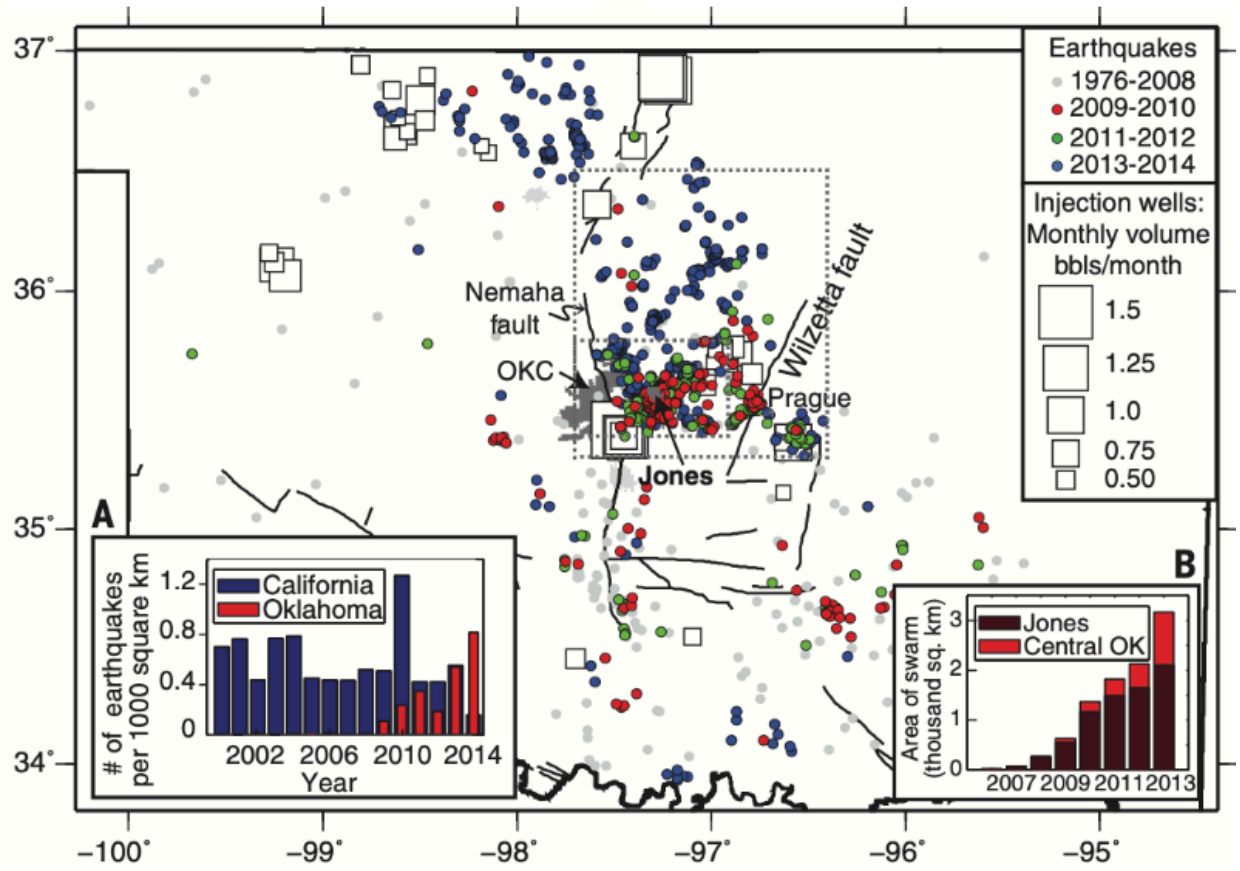


Figure 1.2.3 From Keranen et al., 2014. Large figure shows locations of injection wells and their monthly volume, known major faults, major cities, and seismic events color coded by date range. Subfigure A shows the number of earthquakes per 1000km^2 in California vs. Oklahoma for 2002-2014. Subfigure B shows the area of seismic swarm in both Central OK and Jones for 2007-2013.

CHAPTER 2: METHODOLOGY

2.1 Outline

Catalog data was imported from the following sources using either open-source Python packages or files provided by the OGS: 1) Downloaded from the supplemental materials section of Keranen et al. 2014 as a CSV file, 2) Imported data via ObsPy "Client" from the USGS database, 3) Downloaded data files provided by OGS, and 4) a QuakeML file created via easyQuake (Walter et al., 2020) using the OGS catalog seismograms. The data from these catalogs was imported to a Python working environment, Spyder, for analysis. Files of the catalogs after being relocated with HypoDD (Waldhauser, 2001) were also imported to work with and compared to the original data. Keranen14's catalog was already relocated, so there is not a relocated versus non-relocated analysis set for this catalog. In Python, the event locations are converted from UTM coordinates to a Cartesian coordinate system for calculations and analysis. Events in these catalogs – both relocated and non-relocated – were run through a clustering algorithm and the resulting clusters were then run through an algorithm for linear regression. From these regressions, potential fault segment delineations are identified, and the plots are converted back to UTM coordinates for comparison and presentation of the resulting 2D segments. Length estimates for the segments are controlled by the clustering and regression results, so any limitations or inaccuracies of the prior workflow steps affect length outcome.

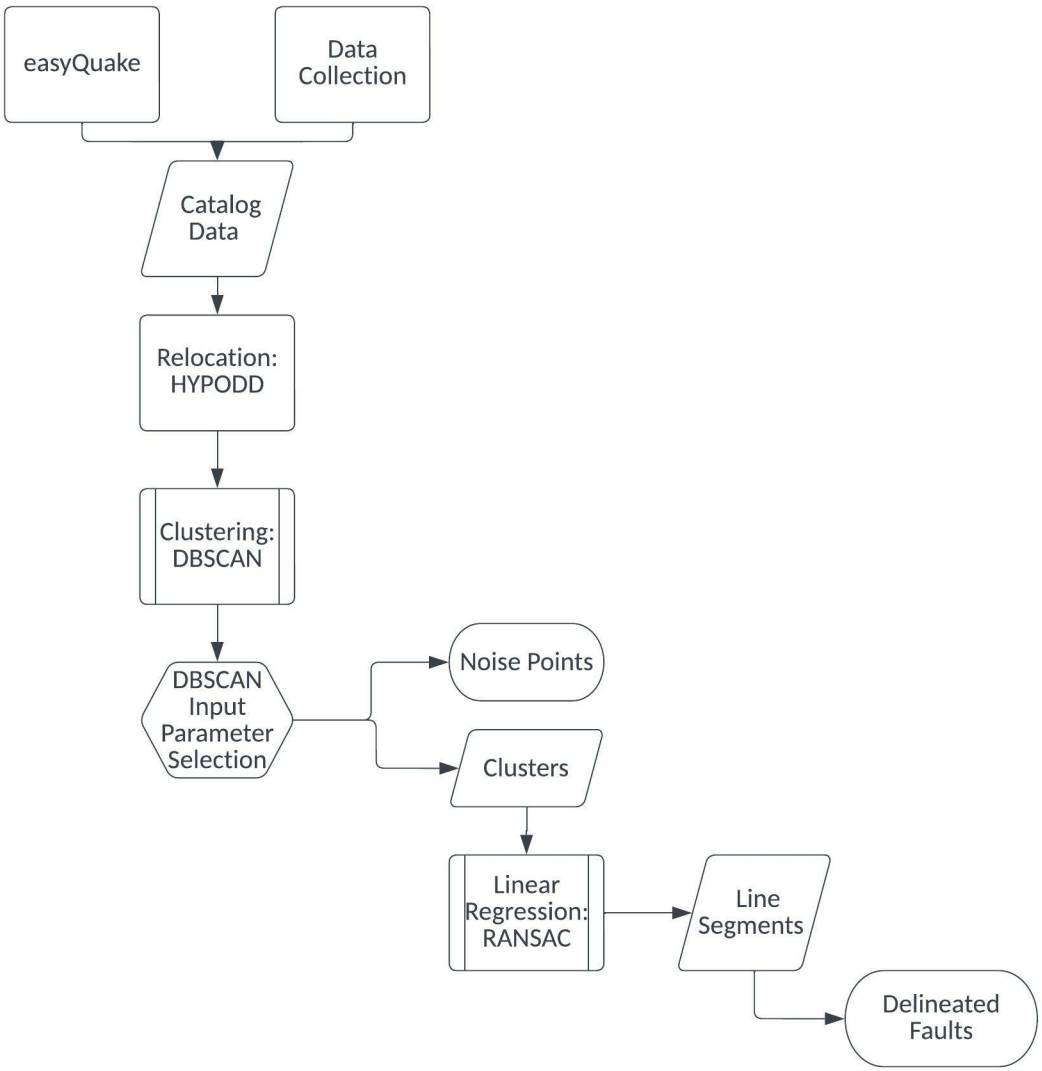


Fig 2.1.1 This figure outlines the workflow for this study. Skoumal et al., 2019's FaultID system for fault identification is adopted as the basis for this study's system. This workflow helps automate fault identification.

2.2 easyQuake

We utilize easyQuake (Walter et al., 2020), a machine learning tool for earthquake catalog development created by researchers at OGS for automated event phase picking and event association. Catalogs created with machine learning tools such as this, rather than by a human manually picking or through template matching, can detect and label phase types for events that are smaller than would otherwise be identified by deterministic algorithms (Walter et al., 2020).

With easyQuake, data that has been previously collected can be used, or current data can be downloaded from International Federation of Digital Seismograph Networks ("FDSN") servers (Walter et al., 2020). It utilizes a deep-learning phase picker, for which there are currently two options from which the user can choose, followed by aggregating the station picks with an adaptation of the PhasePApy 1D associator (Chen and Holland, 2016). Finally, easyQuake calculates local magnitudes and absolute locations from Hypoinverse, then bundles all the associated metadata into a Quake Markup Language, "QuakeML," file (Walter et al., 2021).

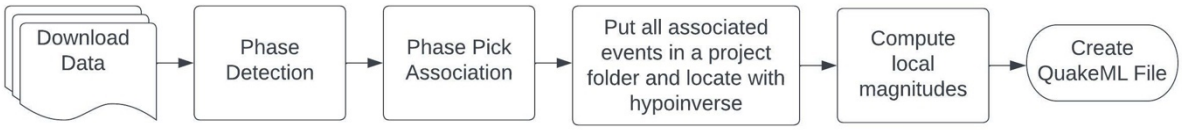


Figure 2.2.2 Workflow for easyQuake. We start with seismic data, move through phase picking and association, then location and magnitude calculations, resulting in an automated file for earthquake analysis. This is a process that is typically done manually by an analyst.

There are currently two ML phase picker options for easyQuake: a generalized phase picker (Ross et al., 2018) or an EQTransformer (Mousavi et al., 2020). Ross et al., 2018's phase picker is a convolutional neural network that scans continuous seismic data and classifies the dominant category of each 4 second window as primarily P-wave, S-wave, or noise. Mousavi et al., 2020's EQTransformer is also a deep learning neural network; it picks first P and S phases based on attention mechanism. Mousavi et al., 2020 notes that not all pieces of a seismic signal are equally relevant for specific classification tasks, so it is useful to determine the relevant sections. The attention mechanism is applied to model the interaction of local (narrow windows around specific phase arrivals) and global (full waveform) seismic features (Mousavi et al., 2020).

Based on early testing of both pickers on Oklahoma data, OGS determined that the pre-trained GPD model performs slightly better than EQTransformer, as it detects more events, so this study utilizes the generalized phase picker option. The phase picker is discussed in further detail in Walter et al., 2020.

After phase picking, the picks are utilized by Chen and Holland's (2016) PhasePApy 1D associator with relevant formatting modifications made to the associator. Upon association, an event magnitude is calculated based on Hutton and Boore (1987) and absolute locations were derived via Hypoinverse.

$$M_L = \log_{10}(A) + 1.11 \log_{10}(D) + 0.00189D - 2.09$$

Equation 1 Hutton and Boore Formulation for local magnitude (M_L) calculation where A is the horizontal displacement maximum amplitude in nm as is denoted in Walter et al., 2020.

Finally, easyQuake aggregates all the pick, amplitude, magnitude, and arrival metadata into QuakeML format. QuakeML files can be converted into hypoDD files for relocation and/or HASH (Hardebeck and Shearer, 2002) input files for determining focal mechanisms. Oklahoma event focal mechanisms have been examined in more detail in past studies (e.g., McNamara, Benz, et al., 2015, McNamara, Rubenstein, et al., 2015, Qin et al., 2019, etc.), so they were not recalculated for this study.

In general, continuous seismograms are located in folders corresponding to a YYYYMMDD format. Location ranges are latitude and longitude based. This study began primarily focused on 2010-2014 data, and a geographic range of roughly 35-36N, -96.5- -98W, following Keranen et al., 2014, (refer to Figure 1.1.2) though was later expanded to cover a broader region. After easyQuake has completed its run, the catalog files may be analyzed as any other catalog file might be, such as relocation and clustering as we have done in this study.

2.3 Relocation

Relocation is commonly conducted after computing absolute locations to account for systematic unmodeled velocity variations in the subsurface. Earthquakes can be relocated with software that uses the "double difference" approach, which is the residual between observed and theoretical travel-time differences (Waldhauser and Ellsworth, 2000). This approach uses an iterative least-squares method to minimize RMS errors for event locations (Waldhauser and Ellsworth, 2000).

One such double-difference relocation software program is HYPODD (Waldhauser, 2001), a FORTRAN package that can be used with event catalogs and/or waveform data from most

seismic networks (Waldhauser and Ellsworth, 2000). When relocated, the relative earthquake locations commonly cluster along linear features that, when combined with focal mechanisms, can constrain the strike and sense of slip of pre-existing faults (e.g., Skoumal, et al., 2019). Each event is related to its neighbors with direct measurements (Waldhauser and Ellsworth, 2000). Figure 2.3.1, below, shows a conceptual illustration of the double difference algorithm.

Two events' initial locations are related to their trial hypocenters based on catalog data or cross-correlation data links. Trial hypocenters can be connected using cross-correlated data or catalog data. The algorithm selects a least-squares solution by finding the vector difference between event hypocenters and makes a series of adjustments through the number of user-selected iterations. Relevant equations for this process are discussed in Waldhauser and Ellsworth, 2000.

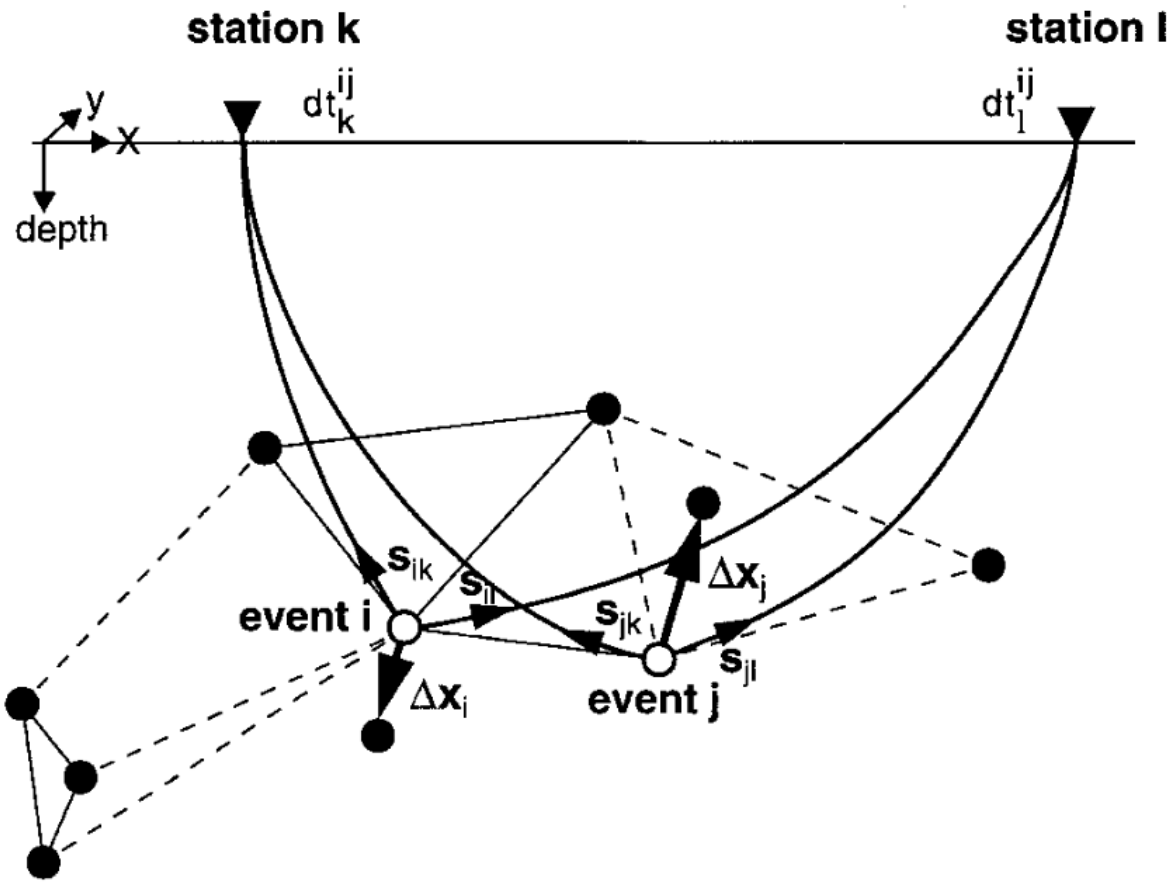


Figure 2.3.1 Waldhauser 2001 Figure 1. An illustration of the double difference relocation algorithm. Circles represent trial hypocenters. These are connected by solid lines for cross-correlation data or dashed lines for catalog data links. The two events, i and j, have initial locations shown with open circles. Stations k and l are shown with inverted triangles above the surface line. The small arrows show the slowness vectors, s, and the larger arrows show the relocation vectors, Δx. dt is the travel-time difference between the two events observed at the stations.

As is described in greater detail in both Waldhauser and Ellsworth (2000) and Waldhauser (2001), the double difference algorithm provides uncertainties an order of magnitude smaller than those for catalog locations. The primary detriment to this system is the availability, or lack thereof, of station data and the way available stations are organized geographically when the event is observed (Waldhauser, 2001). Figure 2.3.2 shows an example of how relocation is used and the differences between a catalog relocation and one done using cross-correlated data.

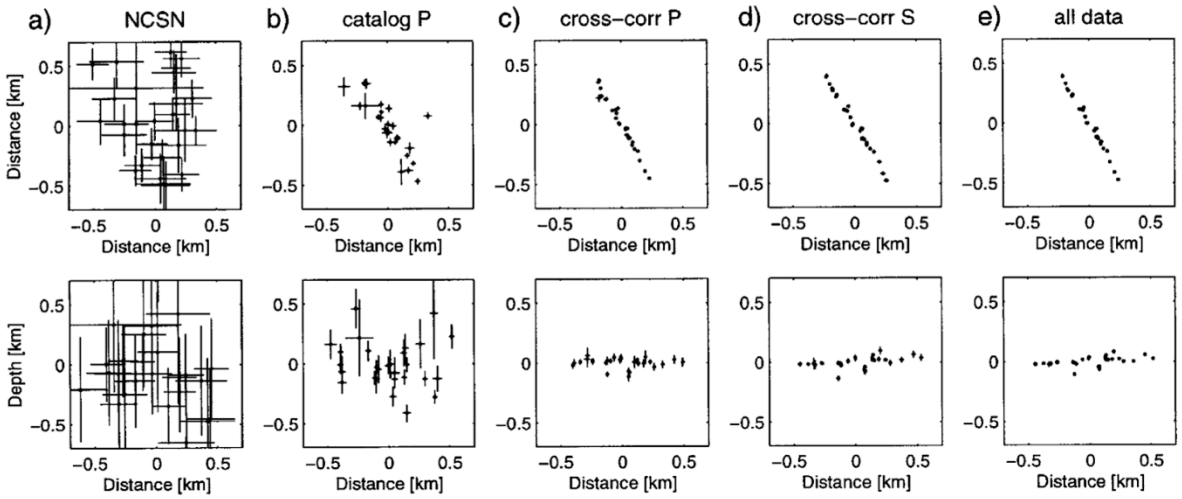


Figure 2.3.2 Figure from Waldhauser 2001. Relocation of California events with 3 independent data sets a. original locations, b. double difference locations, c. P-wave cross correlation times, d. S-wave cross correlation times, and e. all three. Top set is map view, bottom set is NW-SE cross sections). This is an excellent example of how relocating earthquakes can provide a clearer insight to fault lineations.

2.4 Clustering

Events in this study are grouped into clusters based on spatial relationships. Clustering can be carried out by numerous techniques that have, in recent years improved with implementations of machine learning techniques (e.g., Xu and Wunsch, 2005).

After exploring several options for clustering based on previous studies (e.g., Xu and Wunsch, 2005, Skoumal et al., 2019), we chose the Density-Based Spatial Clustering of Applications with Noise, "DBSCAN," algorithm from the scikit-learn Python package (Ester et al., 1996). Clusters are organized based on how densely spaced the events are.

The algorithm is designed to classify arbitrarily shaped data clusters; events are assigned to a cluster if a certain number of points are within a certain radius from other individual events in the "neighborhood" (Ester et al., 1996). This is ideal for working with earthquake data, as events do indeed tend to occur in "arbitrarily" shaped arrangements. While, as previously mentioned, events do tend to cluster along linear features, the edges of these clusters are not uniformly and predictably shaped, so there is a need for the algorithm to be able to account for this arbitrary nature. DBSCAN also allows the user to select some input parameters to better suit the needs of the data.

The user-selected parameters in this study were "eps" and "min_samples." "eps" is the maximum distance – here, in meters – between two events for them to be considered 'in the same neighborhood,' which means this parameter can have a strong effect on which events are sorted into which cluster. The "min_samples" parameter determines how many events must be in a neighborhood for an event to be a 'core point,' and therefore a cluster surrounding that point.

Defining a set of input parameters as “good” is fairly subjective. Visual qualitative analysis is done to inspect clustering results, as well as comparing the number of clusters and number of noise points with the total number of events in the dataset. This can be a lengthy trial-and-error process depending on the data. The final selected input parameters were eps=850m and min_samples=5.

In addition to optimizing input parameters, filters can be used post-clustering to remove errors that are due to other factors. This dataset has filters applied to drop events with the following limits: RMS values less than 0 and greater than 5, vertical errors greater than 3500m, and an azimuthal gap greater than or equal to 180 degrees.

DBSCAN also provides the ability to classify points as ‘noise’ rather than all data points being clustered. This means that earthquakes are not forced into a cluster if there is not one in which it readily fits based on the inputs, which means outliers do not affect the results as greatly (Ester et al., 1996). This ability is beneficial for this study as earthquakes can occur singly, not necessarily as part of a cluster. A clustering algorithm that could not sort out noise points would be more sensitive to outliers and could skew results of cluster lineation orientations (e.g., Dudik et al., 2015).

2.5 Regression

After relocating the events and identifying clusters, mapping is done using regression algorithms to identify linear features. The Random Sample Consensus algorithm, "RANSAC," – again from scikit-learn – is used for this identification. Orientations of clusters' regression lines are what indicate a potential fault segment's presence and orientation. RANSAC was selected over a standard linear regression as it is less sensitive to outliers without introducing a noticeable level of bias beyond that of a standard regression.

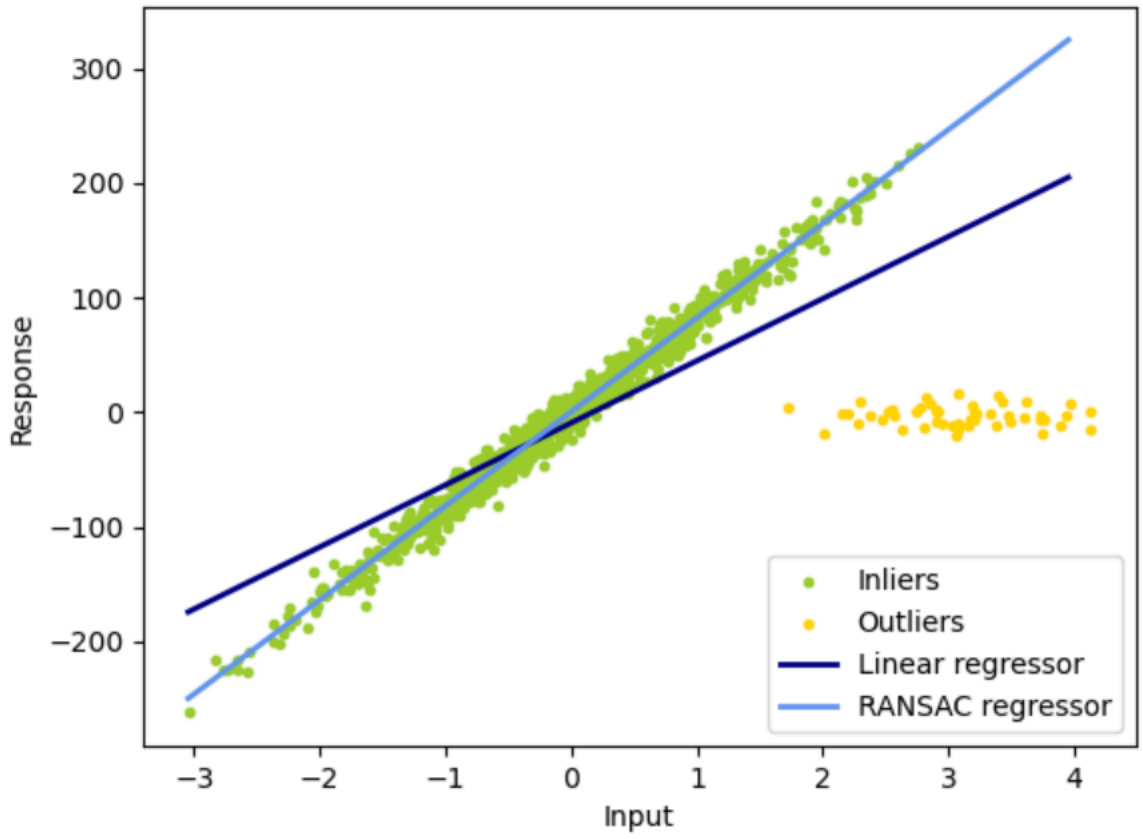


Figure 2.3.3 Figure from https://scikit-learn.org/. Comparison of RANSAC to standard linear regression. Green dots represent inliers of the dataset and yellow dots are outliers. This figure demonstrates how RANSAC creates a regression line more accurately represents the target data. A standard linear regression is more likely to be skewed by outliers than one done using RANSAC.

RANSAC is iterative; unlike standard linear regression RANSAC does not start with the maximum amount of data points and then chip away at outliers, but rather RANSAC uses the smallest feasible initial data set and expands from there to include more inlying data points (Fischler & Bolles, 1981). Much like the ability of DBSCAN to label points as noise, RANSAC's ability to exclude outliers helps keep results more accurate with less manual filtering.

The resulting lineations and corresponding propagation rates of events along them can be used to examine the possibility of future event sequence migrations and increase the available data for earthquake forecasting. Propagation rates of events along these segments are examined via the dates of event occurrences and events' spatial relationships to the segment lineation.

CHAPTER 3: RESULTS

3.1 Catalog Comparisons

The easyQuake catalog captures the most events out of the catalogs examined. When the catalogs are limited by date to work with only 2010 data, and limited geographically to the Jones area, OGS contains 788 and easyQuake contains 2092. The USGS catalog has by far the fewest Oklahoma events reported, with only 130 are in the Jones area. Table 3.1.1 lists the extent of catalog content.

Catalog	Number of events in 2010	Number of events in 2010; limited to Jones region
USGS	193	130
OGS	1053	788
Keranen14	610	586
easyQuake	5336	2092

Table 3.1.1 Number of events reported for each catalog used in this study. This allows for an easy comparison of catalog sizes. The first column lists the catalogs by name. The second column shows the number of events in each catalog for 2010. (NOTE: The USGS catalog is limited here to the Oklahoma region, this is not the entire national catalog) The final column lists the number of events in each catalog for 2010 in only the Jones region.

The Gutenberg-Richter plots shown in Figures 3.1.1 through 3.1.6 are based on the non-relocated catalogs, as relocation does not change the magnitude. The non-relocated data contains events that were unable to be relocated, though they occurred.

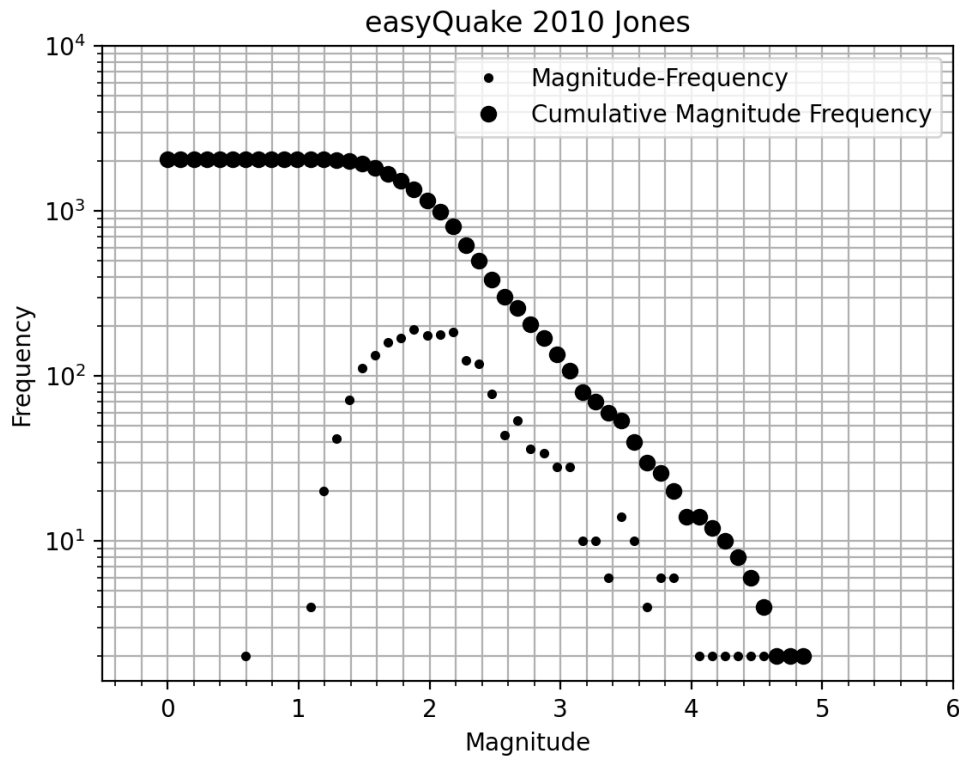


Figure 3.1.1 A Gutenberg-Richter plot of the easyQuake catalog of events in the Jones area for 2010. The large dots are Cumulative Magnitude Frequency of events. The small dots are Magnitude-Frequency.

Based on the plot in Figure 3.1.1, above, the magnitude of completeness for the easyQuake 2010 Jones data is approximately M1.6 as this is where the relationship of the event cumulative magnitude frequency and event magnitude values stops behaving in a logarithmically linear fashion. This is a visual estimation based on the resulting plots.

The easyQuake catalog appears to have the smallest magnitude of completeness of the group.

Visual estimates for the magnitudes of completeness are shown to be 2.2 for OGS, 2.5 for USGS, and 2.0 for easyQuake.

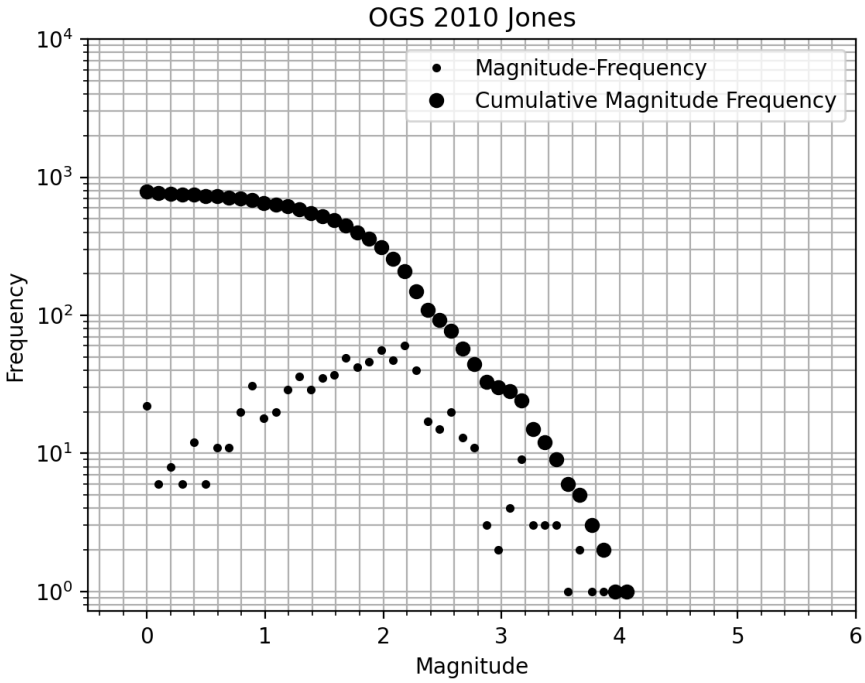


Figure 3.1.2 A Gutenberg-Richter plot of the OGS catalog of events for 2010 in Jones, OK.

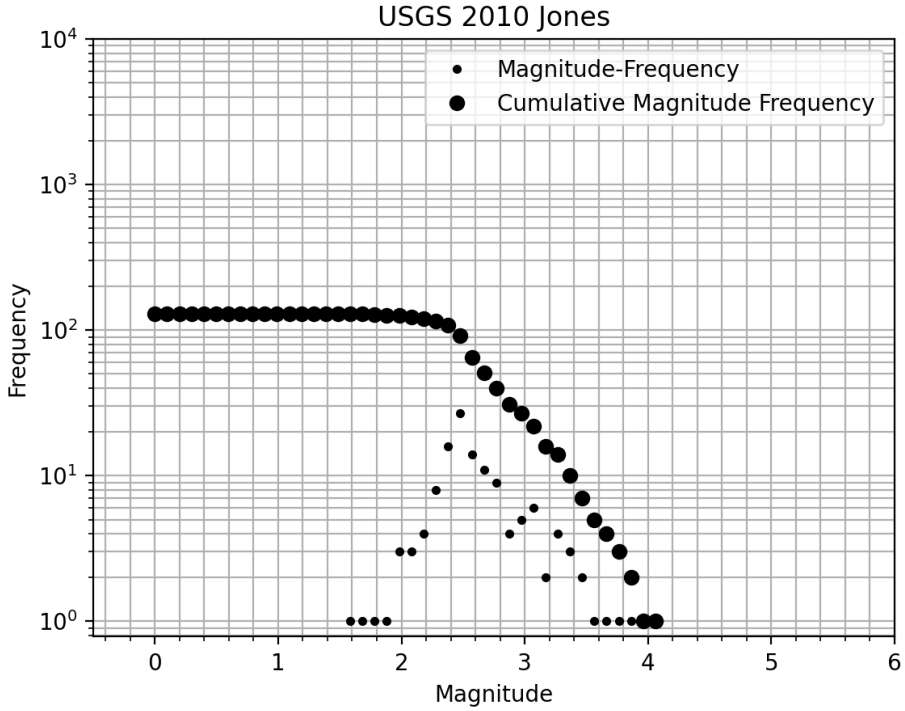


Figure 3.1.3 A Gutenberg-Richter plot of the USGS catalog of events for 2010 in Jones, OK.

Examining the entire 2010 catalogs, rather than only the Jones area, provides a clearer picture of the events' magnitude and frequency relationships that year. The Jones-only data still shows the same general relationship, but with less clarity and lower accuracy due to the small size of the datasets. Figures 3.1.4 through 3.1.6 show Gutenberg-Richter plots for each of the three catalogs during 2010 with no geographic limitations.

Though based on larger sample sizes, the relationships stay the same, with easyQuake containing the most events, and USGS containing the fewest. For 2010, the OGS catalog contains 1053 events, while the easyQuake 2010 catalog contains 5336. The USGS reports 193 Oklahoma events in 2010. This is not the entirety of the USGS 2010 catalog; it is limited to the Oklahoma region but covers the entire state. Keranen14's catalog contains 610 events in 2010, and 586 within the geographic limitations. Gutenberg-Richter plots were not created for Keranen's catalog as magnitude data was not available.

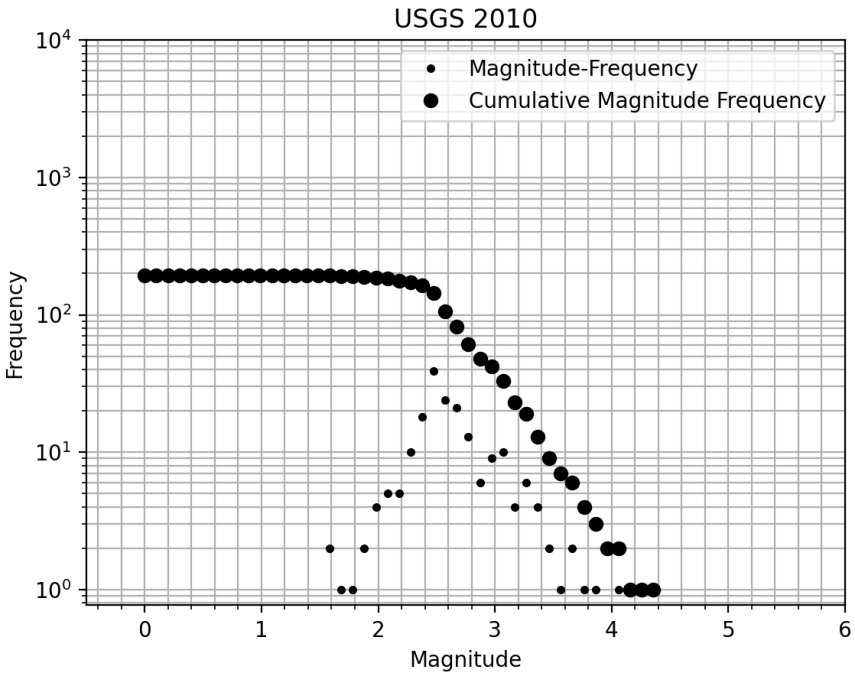


Figure 3.1.4 A Gutenberg-Richter plot of the USGS catalog of events for 2010.

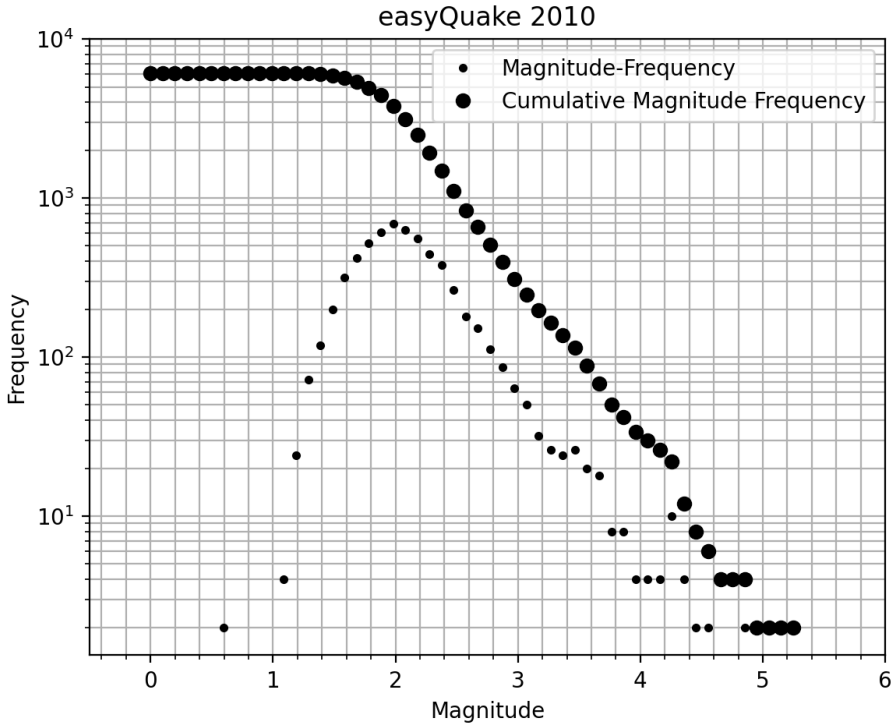


Figure 3.1.5 A Gutenberg-Richter plot of the easyQuake catalog of events for 2010.

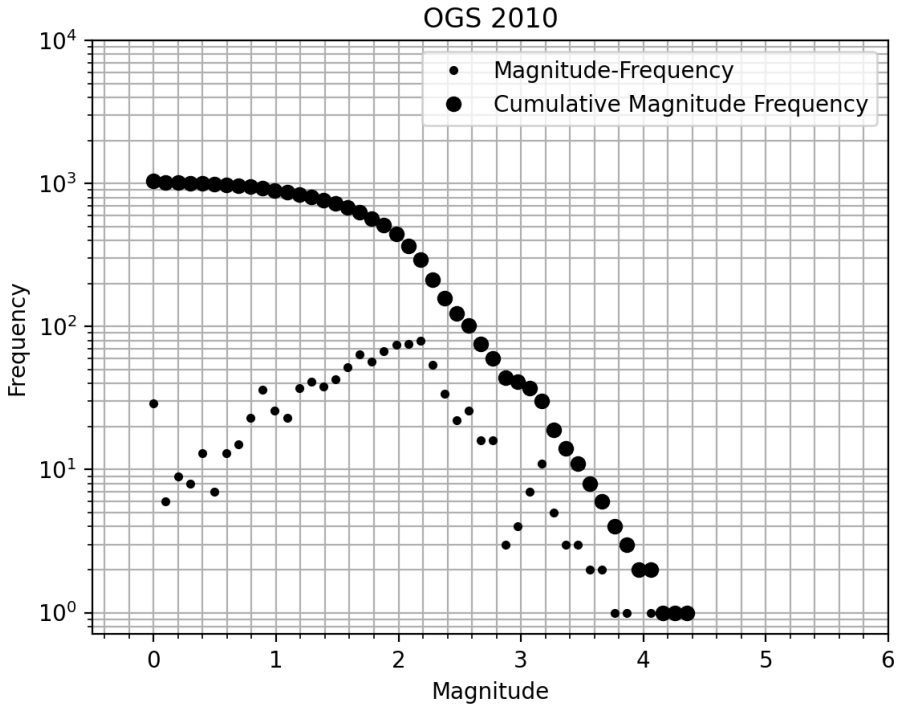


Figure 3.1.6 A Gutenberg-Richter plot of the OGS catalog of events for 2010.

3.2 Original Data Versus Relocations

We created plots of both the relocated and non-relocated earthquake catalogs for the same region. Figures 3.2.1 through 3.2.3 show these plots for the easyQuake 2010 data. Figure 3.2.1 displays the original, non-relocated event locations, while figures 3.2.2 and 3.2.3 display relocated events.

Events are assigned a color based on the date of occurrence on a gradient scale, with older events in purple(oldest) to blue shades, and newer events in green to yellow(youngest). Earliest events are shown to occur in late March, and latest events occur in mid-to-late December.

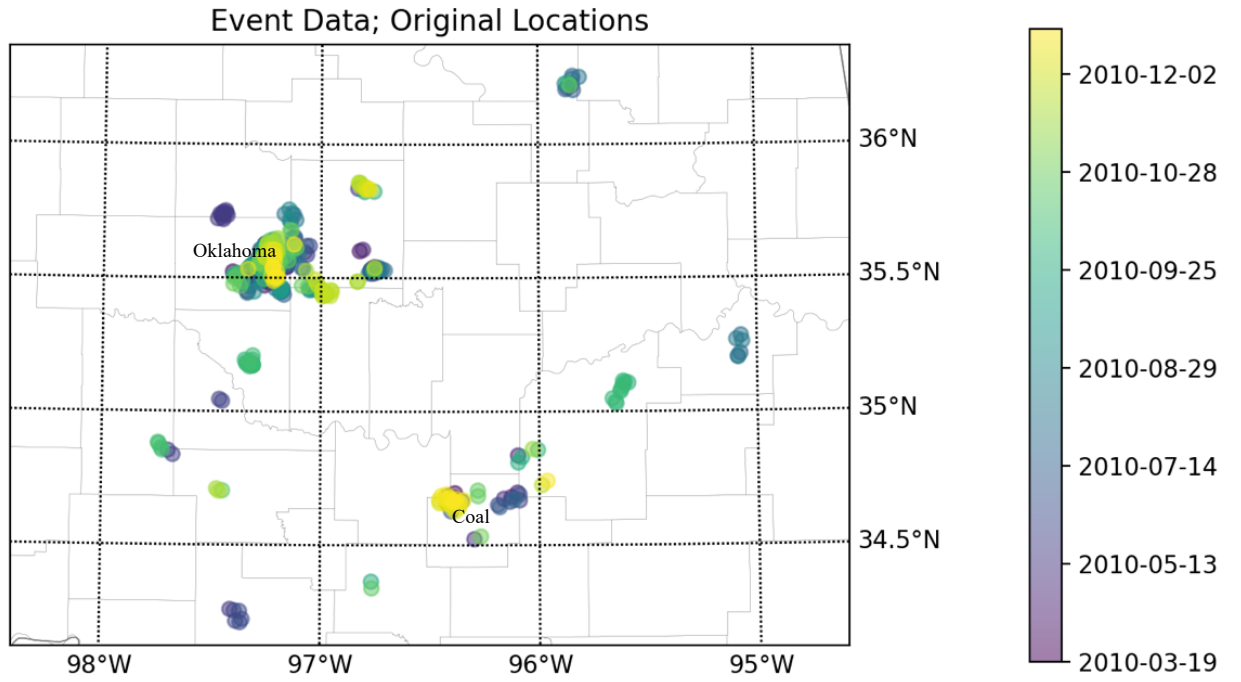


Figure 3.2.1 Original (non-relocated) locations for the easyQuake 2010 data

The most seismically active area is the eastern side of Oklahoma County, just northwest of 35.5°N, -97°W. Jones, OK falls within this region. This area is shown to have events

throughout the year. To the southeast, Coal County is also fairly active, but with fewer events than Oklahoma Co. There are also groups of events that happened at relatively the same time as each other in several other counties, though these counties do not show year-round activity like Oklahoma and Coal counties do.

Relocation of the data provides a more accurate picture of the event locations. Even with the default parameters for HYPODD relocation (Figure 3.2.2), the events are more tightly arranged than before any relocation. However, too many valid events are dropped in this case, so the input parameters need to be adjusted to minimize this issue.

Figure 3.2.3 shows the 2010 easyQuake data using the optimized relocation parameters. It has more events successfully relocated than Figure 3.2.2, while remaining more tightly organized than Figure 3.2.1. These optimized parameters are used for relocation of all catalogs, not for only the 2010 data. It allows for as many valid events as possible to be relocated.

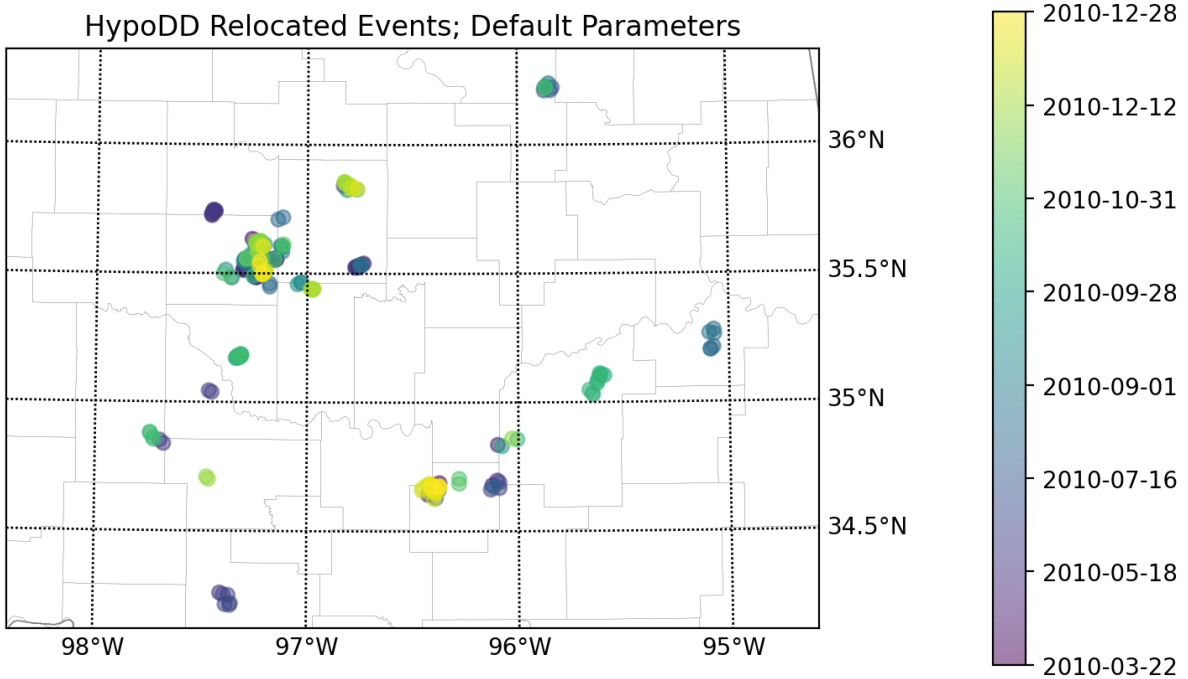


Figure 3.2.2 easyQuake 2010 data relocated via HYPODD with the default parameters

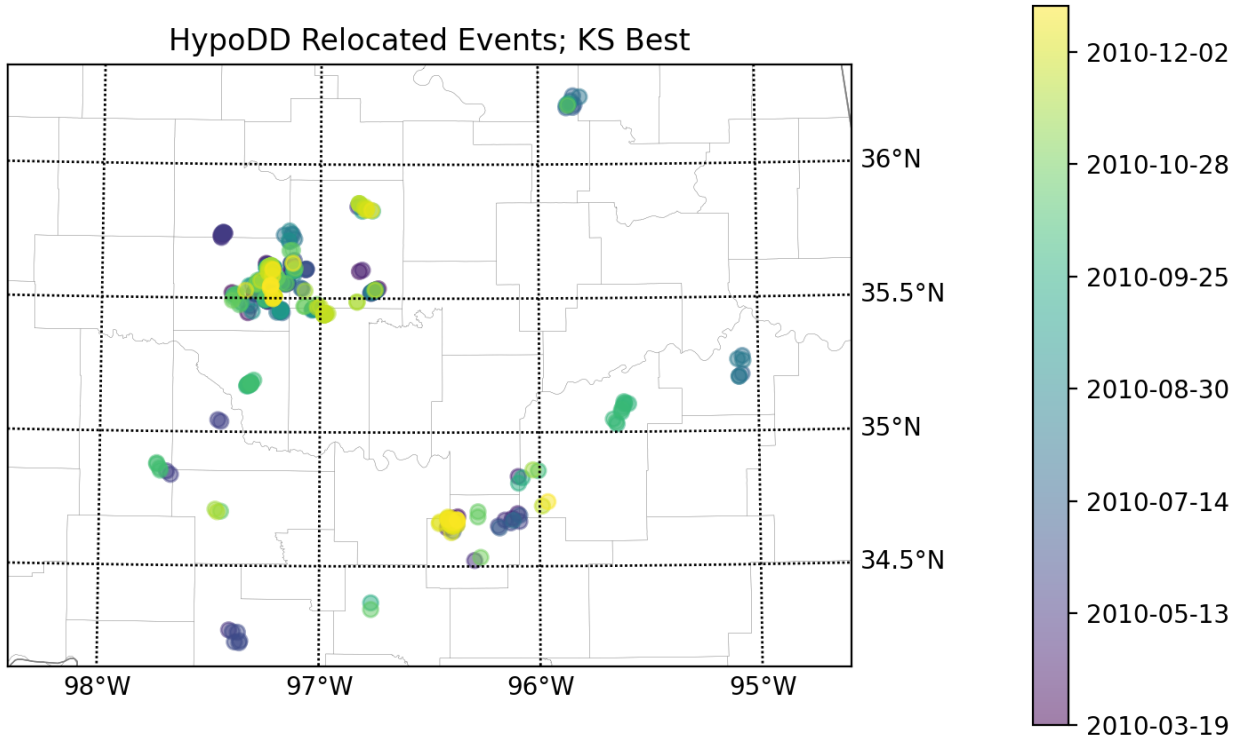


Figure 3.2.3 easyQuake 2010 data relocated via HYPODD with optimized, user-selected parameters, labeled "KS Best"

After catalogs have been relocated, they can be more appropriately compared. Figure 3.2.4 shows a closer look at how relocated data is more tightly organized than before relocation, which assists in more accurately organizing events into clusters. More accurate clusters mean more accurate linear regressions, which in turn means more accurate fault segment delineations. The optimized parameters are user-defined and are not easy to determine, though once a parameter set has been established for a region it is easy to apply. It is important for the user to spend time researching how HYPODD works and how the parameters affect the relocations. Making random adjustments does not do the trick – one needs to actively work toward building an intuition for what the data needs for a proper relocation.

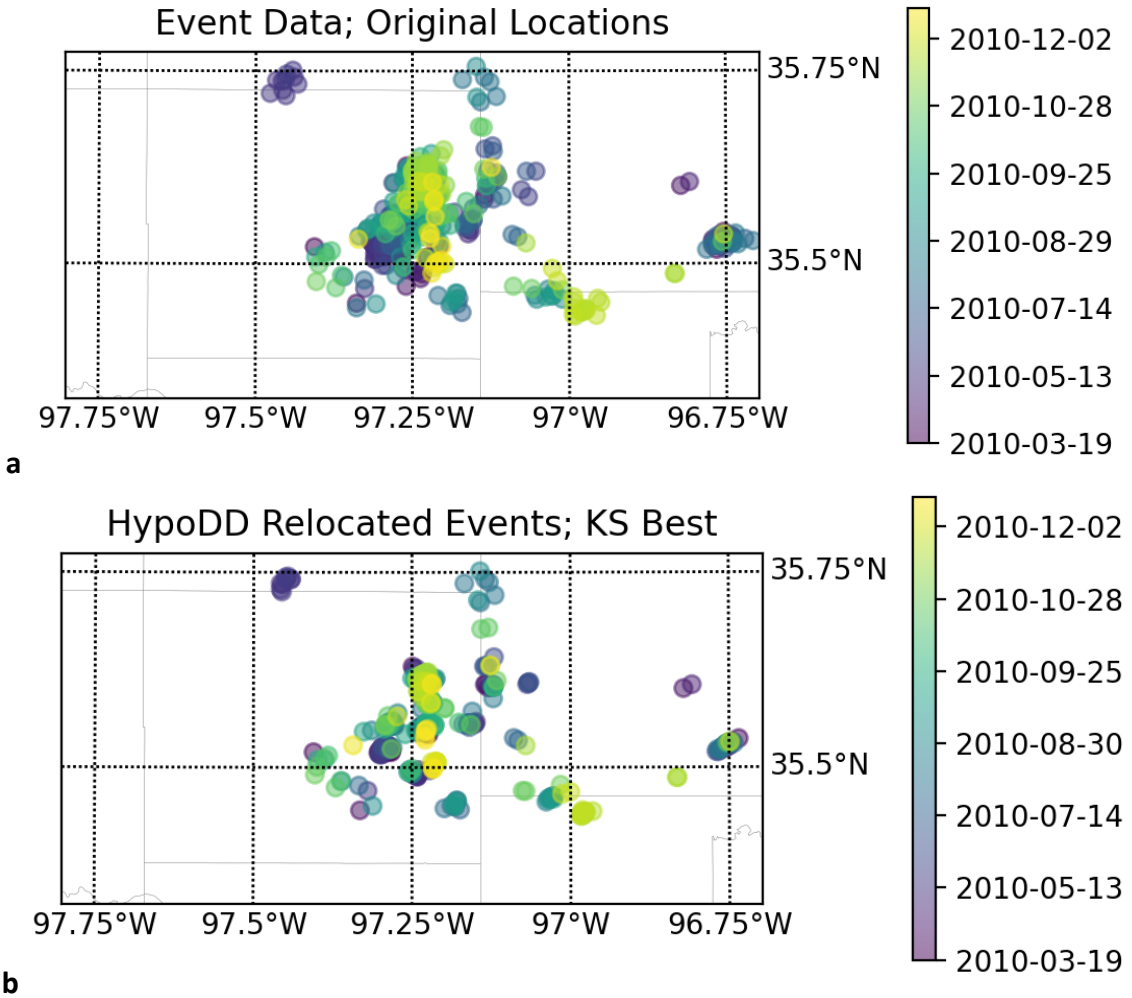


Figure 3.2.4 (a)easyQuake data for 2010, non-relocated. (b)easyQuake 2010 data relocated via HYPODD with optimized, user-selected parameters, labeled “KS Best”. These show us how relocated data is more tightly organized than before relocation, which assists in more accurately organizing events into clusters. More accurate clusters = more accurate linear regressions = more accurate fault segment delineations.

3.3 Comparison to Keranen et al., 2014

As seen in Figure 1.1.2 c, Keranen14 created a plot of the distance from the injection well at approximately 35.43°N, -97.46°W. Figure 3.3.1 is a recreation of this plot using the catalog data provided in the supplementary materials section of Keranen et al., 2014. This plot presents a ramp-like structure, inclining from March to October (right to left). From March to June, most events are roughly 18km from the well. Event-to-well distance steadily increases to approximately 25km June through September, then plateaus near 27km from September to October.

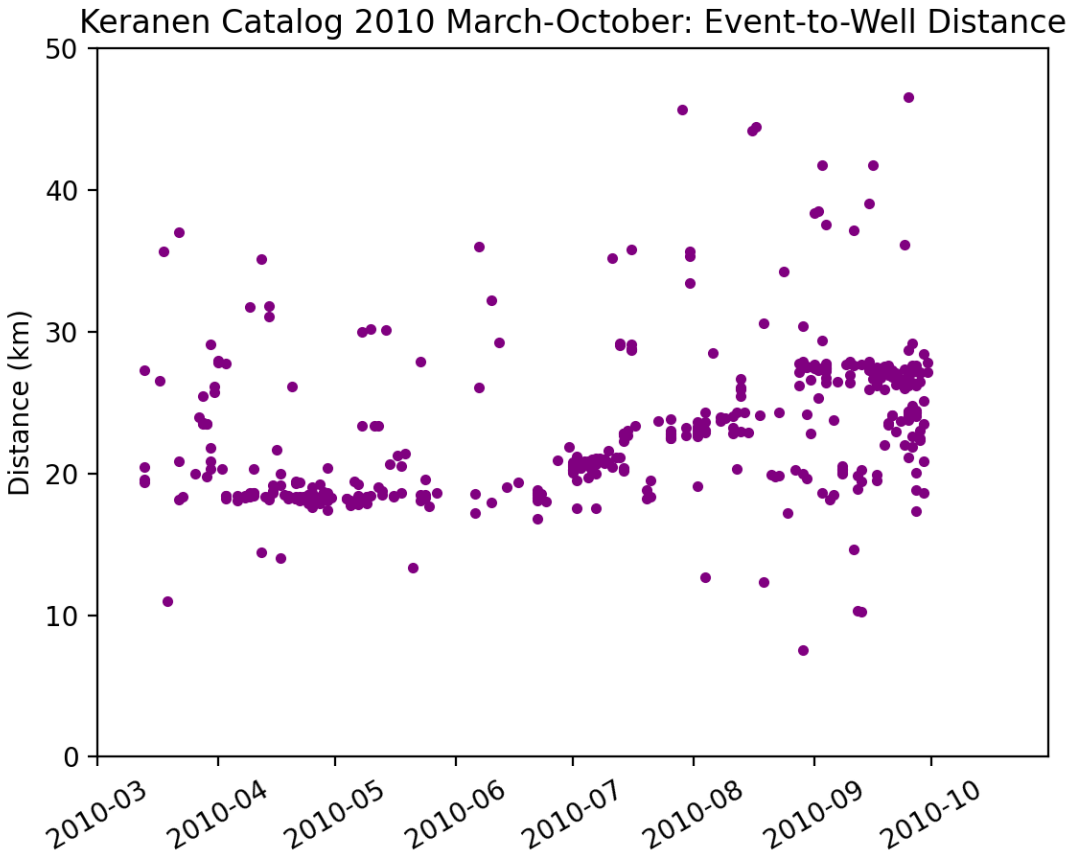


Figure 3.3.1 Recreation of Figure 1.1.2c using the Keranen et al., 2014 catalog for March through October of 2010. This plot prominently displays a ramp-like feature that indicates possible event migration away from the high-volume wastewater injection site.

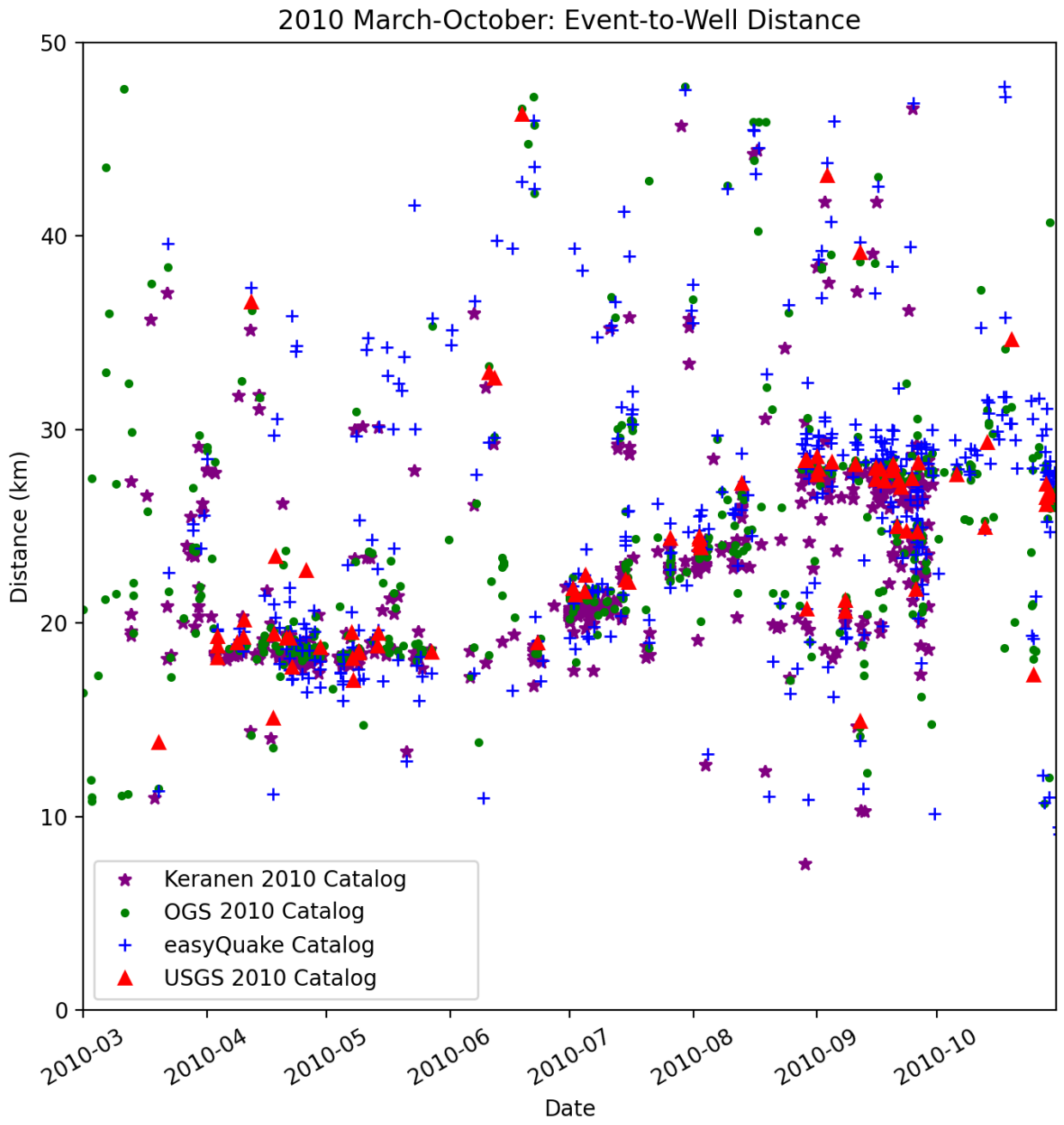


Figure 3.3.2 Keranen, OGS, easyQuake, and USGS data for March through October of 2010. Distances from events to the injection well site at 35.43°N, -97.46°W. Limited to a maximum of 50km from the injection site. Stars (purple) represent Keranen events. Dots (green) represent OGS events. Crosses (blue) represent easyQuake events. Triangles (red) represent USGS events. We see here that the ramp is not an artifact of the Keranen14 data but appears in all four of the examined catalogs.

When examining all four catalogs simultaneously on the same scale as Figure 1.1.2 c, as is displayed in Figure 3.3.2, it can be seen that the “ramp” does not appear to be an artifact of Keranen14’s catalog data, though the feature is less apparent when utilizing the easyQuake catalog. This is more clearly shown below, in Figure 3.3.3. All four catalogs contain the same structure at roughly the same location. They also all show a set of events throughout the zone from approximately 15km to 30km all occurring in October 2010.

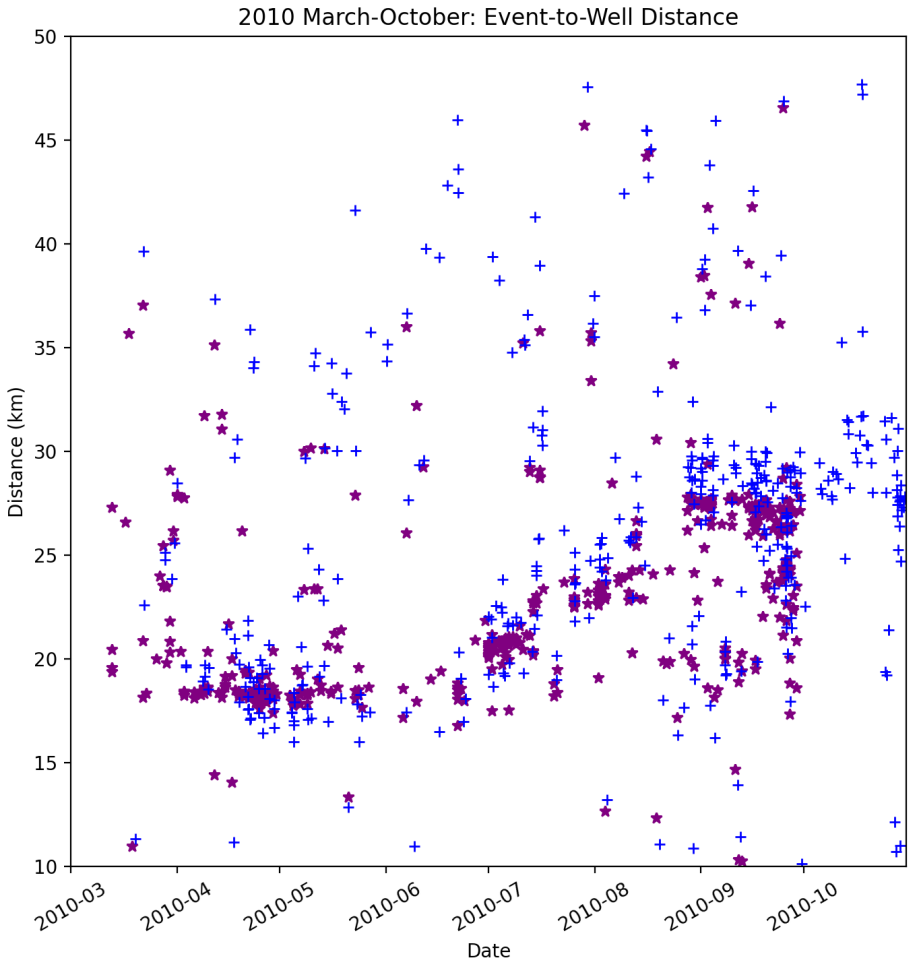


Figure 3.3.4 Keranen14(purple stars) and relocated easyQuake(blue crosses) data for March through October of 2010. This allows us to see the differences more clearly between the ramp structure as seen in the Keranen14 and easyQuake catalogs. We see that the ramp is less apparent of a feature in the easyQuake than the Keranen14

However, when examining the rest of the catalog, which includes data further away, the ramp is no longer a prominent structure. The farthest event from the well in the Keranen14 catalog for this date range is roughly 150km away. Though these events are included in the catalog, they are far enough away from the wells such that they may not be related to the well activity.

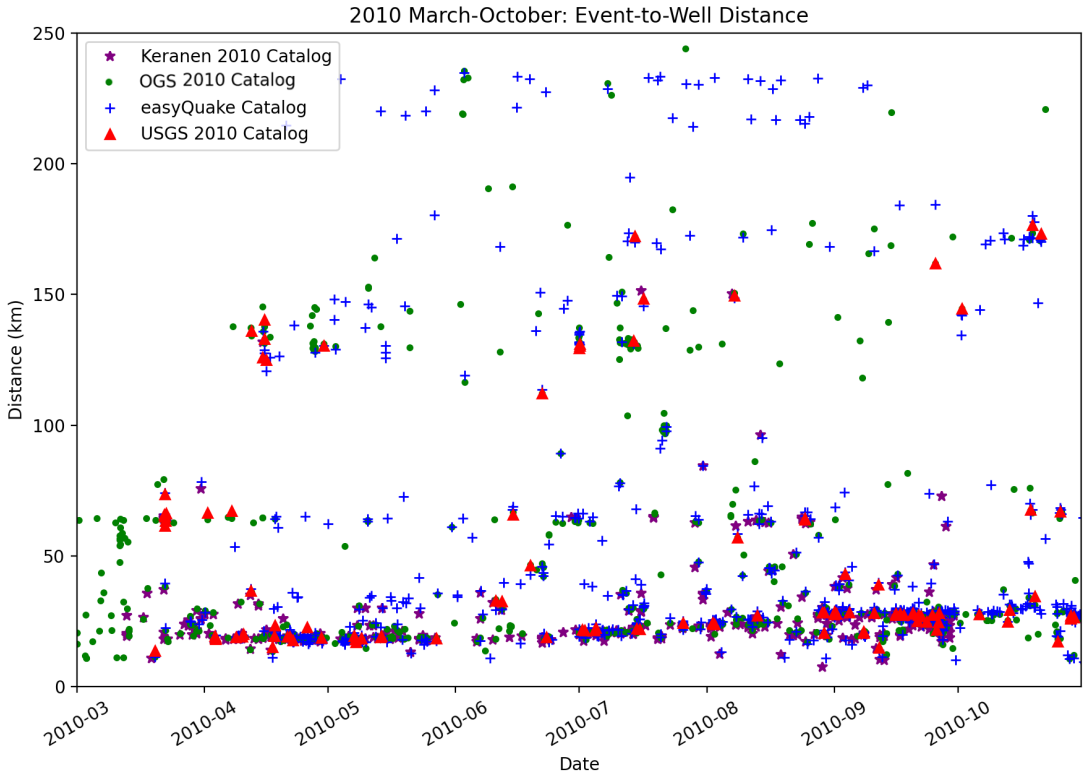


Figure 3.3.5 Keranen, OGS, easyQuake, and USGS data for March through October of 2010. Distances from events to the injection well site at 35.43°N, -97.46°W. Stars (purple) represent Keranen events. Dots (green) represent OGS events. Crosses (blue) represent easyQuake events. Triangles (red) represent USGS events.

3.4 Lake Levels

Some cases of seismicity have been reportedly induced by hydraulic loading pressure changes in places such as France, Canada, China, and India (e.g., Costain & Bollinger, 2010, Foulger et al., 2017, Grasso et al., 2018).

In addition to the comparisons of Keranen's data to other catalogs, reservoir data was examined in the Jones area to determine the area's hydraulic loading levels' contribution to the increased seismicity. Figure 3.4.1 shows the reservoir locations relative to Jones, and Figure 3.4.2 compares the gage height levels with local seismicity.

Two reservoirs in the vicinity are Wes Watkins Reservoir, shown in red, and the Shawnee Reservoir, in blue. While the Wes Watkins Reservoir gage height is consistently higher than that of the Shawnee Reservoir, the two reservoirs' levels' movements mimic each other. Gage heights are at a local maximum during 2010, and generally decrease until reaching a low point in 2013. After this low point there is a spike in both reservoirs, before steadily decreasing again.

Seismicity rates in this area as reported by the USGS do not follow the same trend as the reservoir gage heights. The seismicity for the area is fairly consistent, topping out near five earthquakes greater than M2.5 per day (Figure 3.4.1a) and staying below fifty per month (Figure 3.4.1b) for 2009 through 2013, with a sharp increase beginning sometime between 2013 and 2014. The beginning of said sharp increase does happen to coincide with the spike in both reservoirs' gage heights, but as the gage heights return to decreasing, the seismicity rate continues to increase.

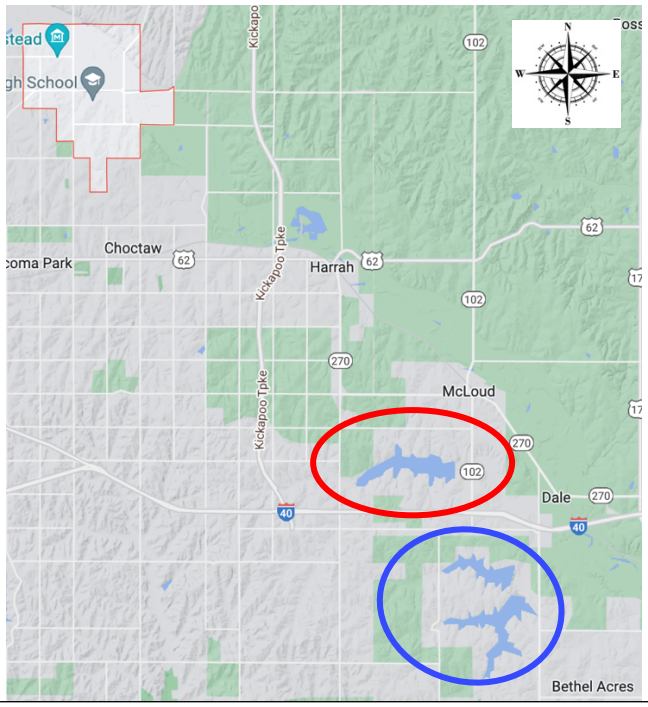


Figure 3.4.1 Map view of reservoirs with relation to Jones, OK. Jones is positioned in the upper left (Northwestern) corner of this map. Wes Watkins Reservoir is circled in red and Shawnee Reservoir circled in blue. Lake Stanley Draper is also in the vicinity but not shown in this map. It falls just outside the map in the southeast corner. Lake level data was not readily available at the time of study for Stanley Draper, so only Wes Watkins and Shawnee are evaluated here.

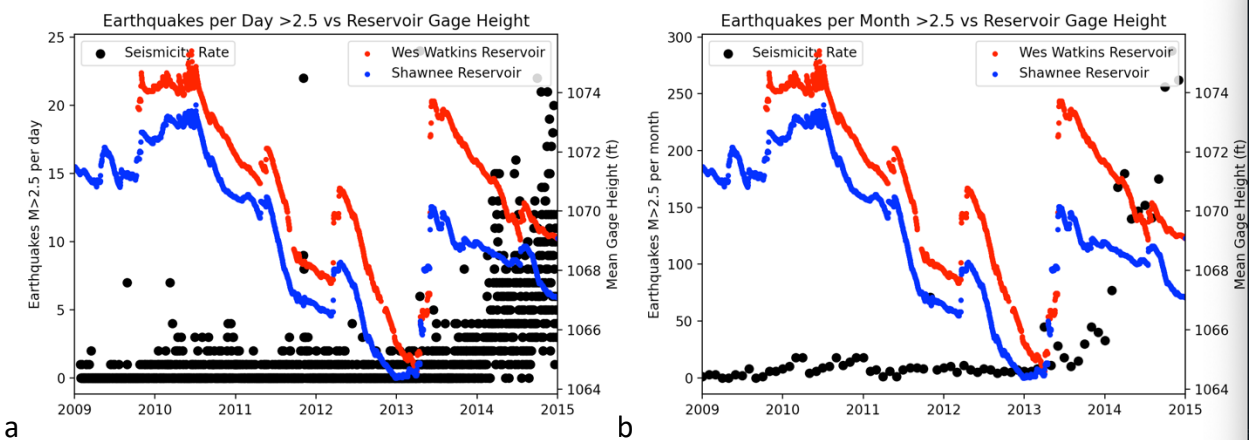


Figure 3.4.2 Plots of local seismicity (black) versus reservoir gage heights (red and blue) for 2009-2015. (a)Earthquakes larger than M2.5 per day vs gage height. (b) Earthquakes larger than M2.5 per month vs gage height. The apparent lack of correlation between gage heights and seismicity rates indicates that these lake levels are not likely tied to the increase in earthquake occurrences during this timeframe.

3.5 Clusters and Lineations

For each catalog examined, a figure was created to look at both the original and relocated datasets. Figure 3.5.1 below, demonstrates the differences between clustering results for relocated and non-relocated data, with the exception of the Keranen14 catalog, as it was previously relocated for the Keranen14 study before distribution.

Clusters are determined by DBSCAN. Tables 3.5.1 through 3.5.4 compare resulting clusters and noise points for each step in the trial-and-error process for optimizing input parameters. Selected parameter runs are highlighted green. Tables are listed in the order for which parameters were tested.

The final parameter selections, as highlighted in the tables, are as follows: eps=850, min_samples=5, event dropped if vertical error above 3500, and event dropped if RMS below 0 or above 5.

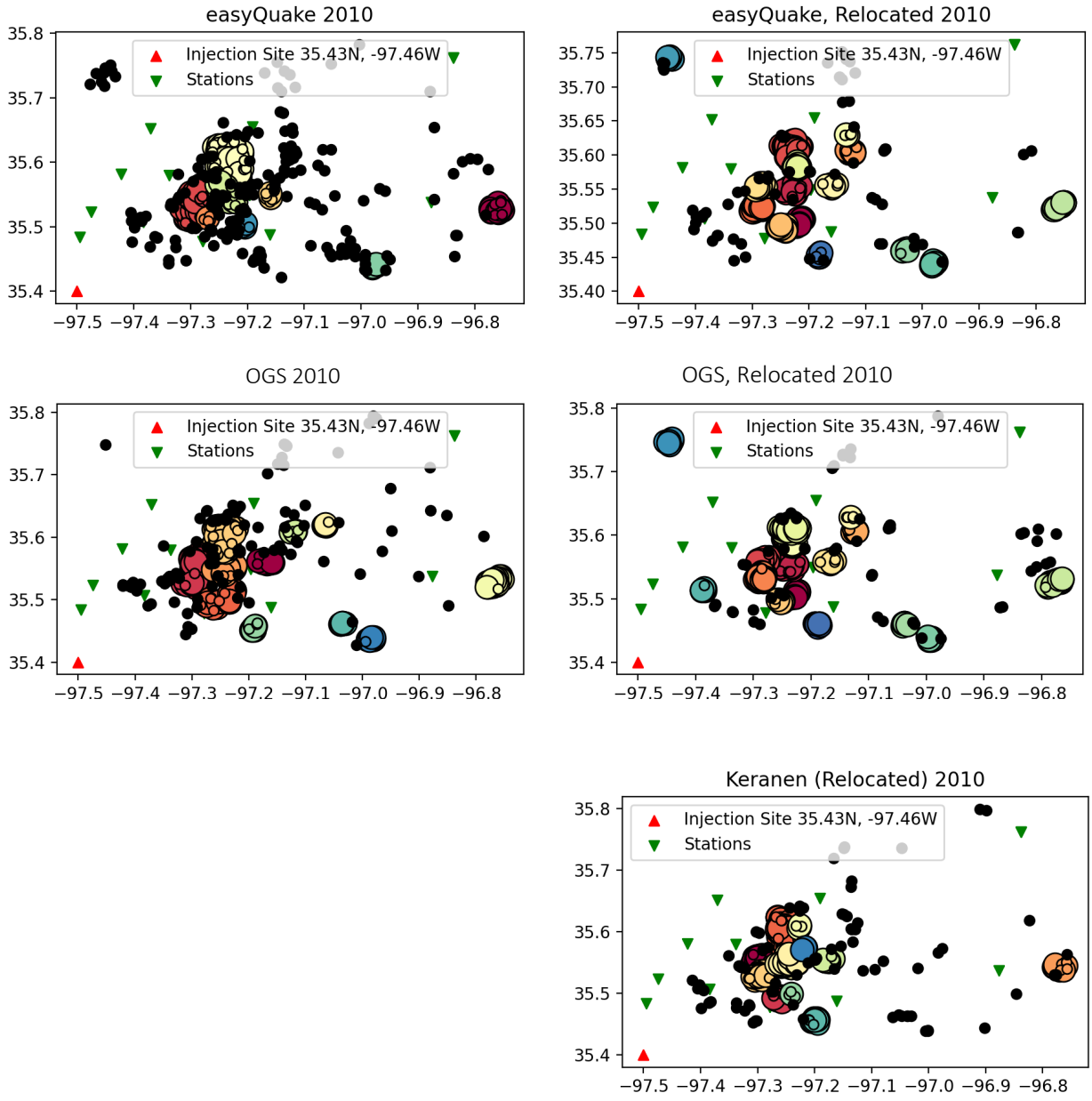


Figure 3.5.1 Keranen14, OGS, easyQuake, and USGS data for 2010. Red Triangle represents injection well site. Green, Inverted Triangle represents seismic monitoring stations. Black dots are events that are considered noise points and are not part of any cluster. Circles with color indicated clusters.

eps	min_samples	Clusters	Noise Points
800	5	1412	22519
700	5	1693	26749
750	5	1547	24583
775	5	1499	23479
900	5	1246	19342
900	7	845	24545
1000	7	747	21269
600	7	1360	41873
800	10	675	36333
1200	10	450	21052
1300	10	425	18990
1300	15	282	25656
1300	5	872	12313
1500	5	742	10327
700	7	1108	34345
700	6	1511	29376
600	6	1776	36089
600	5	2035	32631
650	5	1826	29447

Table 3.5.1 DBSCAN Trial values. Trials to select value for min_samples. Final selection highlighted in green.

n	eps	min_samples	Clusters	Noise Points
5000	800	5	1476	22093
5000	700	5	1774	26229
5000	600	5	2050	31967
5000	750	5	1594	24118
5000	850	5	1400	20407
5000	900	5	1300	18927

Table 3.5.2 DBSCAN eps Trials with a Vertical Error filter. Vertical Error greater than n is dropped. Trials done to select value for eps. Final selection highlighted in green.

n	Clusters	Noise Points
5000	1400	20407
2500	1408	21093
1000	1342	21405
10000	1293	19500
7500	1345	19944
6000	1384	20207
5500	1397	20334
5250	1394	20384
4500	1402	20514
3000	1411	20878
3500	1418	20710

Table 3.5.3 Vertical Error trials. eps = 850. min_samples = 5. Events are dropped if the Vertical Error is greater than n. Trials done to select value for Vertical Error (here, n). Final selection highlighted in green.

RMS	Clusters	Noise Points
No Filter	1848	26950
RMS<1	1207	16855
RMS>0	1631	24584
1>RMS>0	1196	16855
10>RMS>0	1631	24584
8>RMS>0	1549	23358
5>RMS>0	1412	20668
2>RMS>0	1299	18431
3>RMS>0	1371	19362

Table 3.5.4 Trials to select RMS filter. eps = 850. min_samples = 5. Events are dropped if the Vertical Error is greater than 3500.

The two figures below (Figures 3.5.2 and 3.5.3) show delineated clusters for the easyQuake 2010 data. These figures demonstrate the difference relocation makes on both clustering and delineation results. The relocated data has more events sorted into clusters and fewer noise points. This also allows the linear regressions to reflect the fault segments and their orientations more accurately, as the clustering is more accurate.

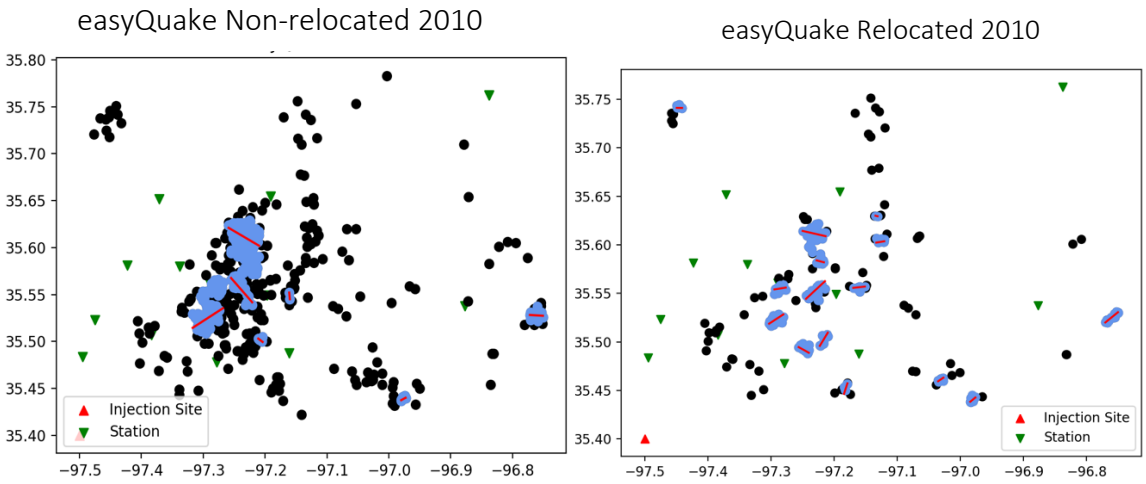


Figure 3.5.2(left) Plot of linear regressions for easyQuake 2010, non-relocated events.

Figure 3.5.3(right) Plot of linear regressions for easyQuake 2010, relocated events.

In Figure 3.5.4, we see the relocated easyQuake catalog for 2010 after clustering, both before and after regression. Figure 3.5.4 shows the data after clustering, both before(3.5.4a) and after(3.5.4b) delineation. Clustering for the 2010-2021 catalogs use the above-defined parameter selections. The individual year catalogs have an eps of 800m, min_samples of 5, and are not filtered further.

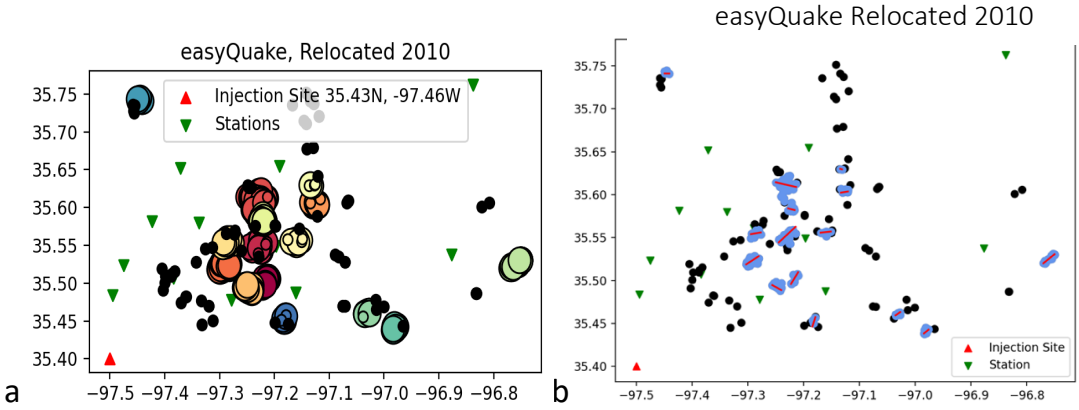


Figure 3.5.4 (a) Cluster results for easyQuake 2010 relocated catalog in Jones. Black dots represent noise points. Colored circles show events belonging to clusters (b) Delineations of clusters for easyQuake 2010 relocated catalog in Jones. Blue dots represent events belonging to clusters. Black dots represent noise point events. Red lines show regressions.

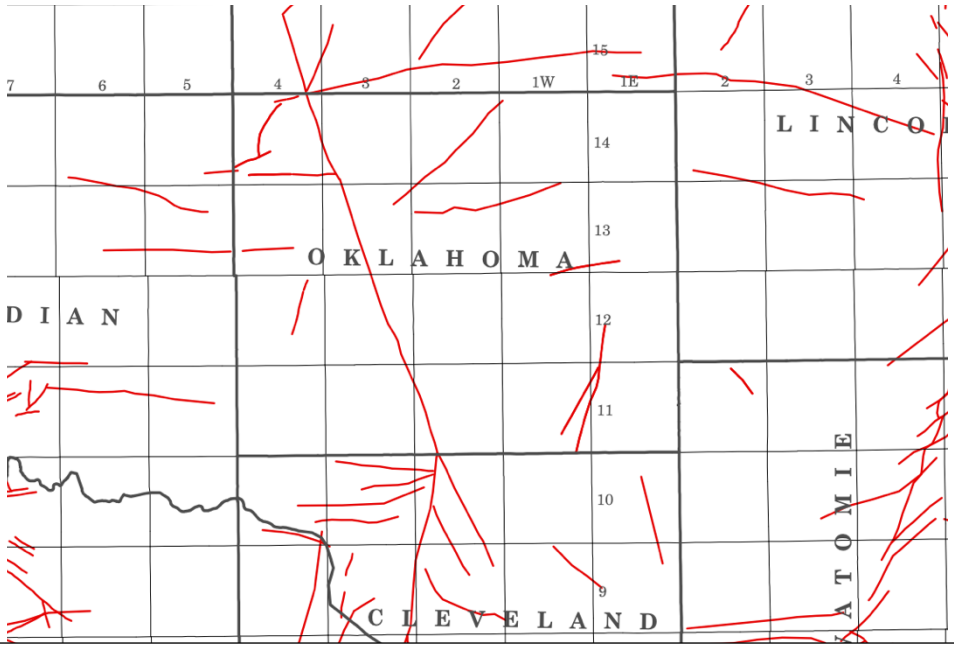
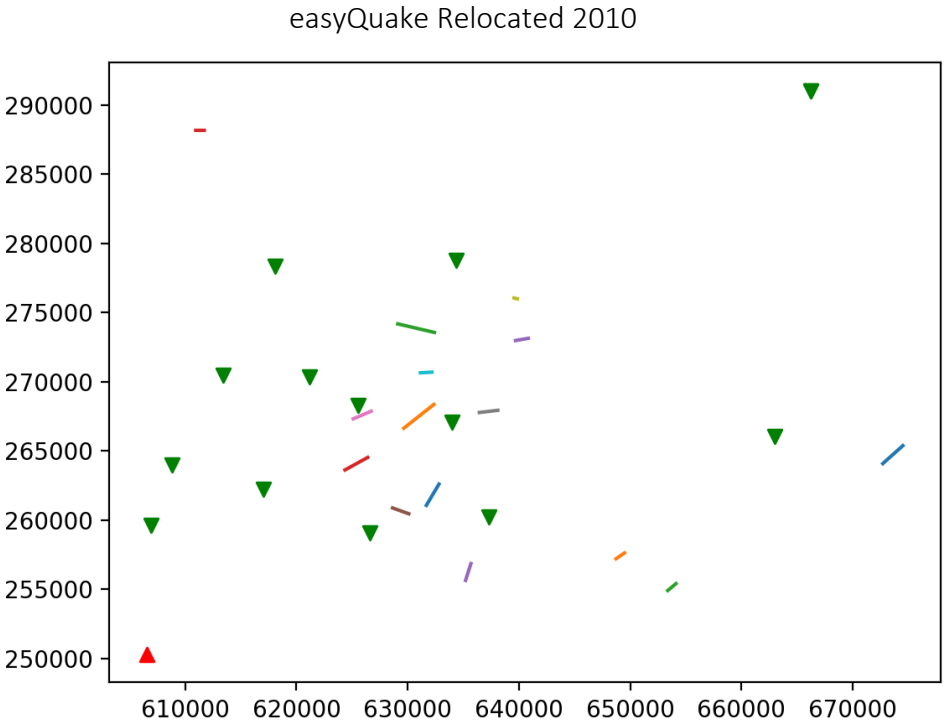


Figure 3.5.5 This is a portion of the current official preliminary fault map of Oklahoma as reported by the USGS. It was created by researchers at OGS and was published in 2015. It is clear that Oklahoma county does not have many mapped fault segments as represented in the official map. One of the main goals of this study is to help fill the gaps of unmapped seismogenic fault segments.



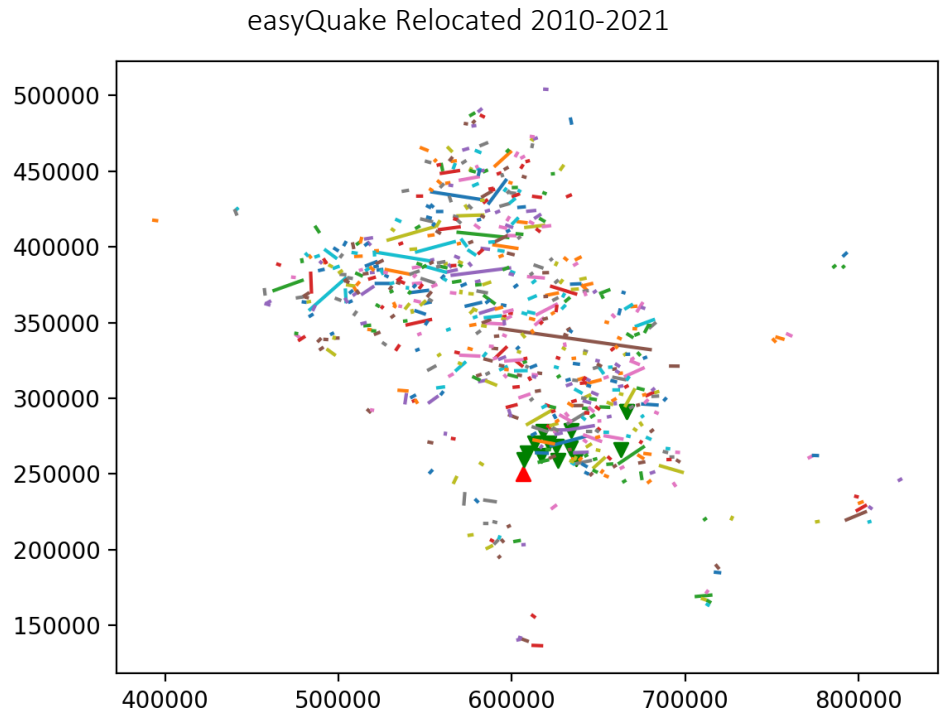
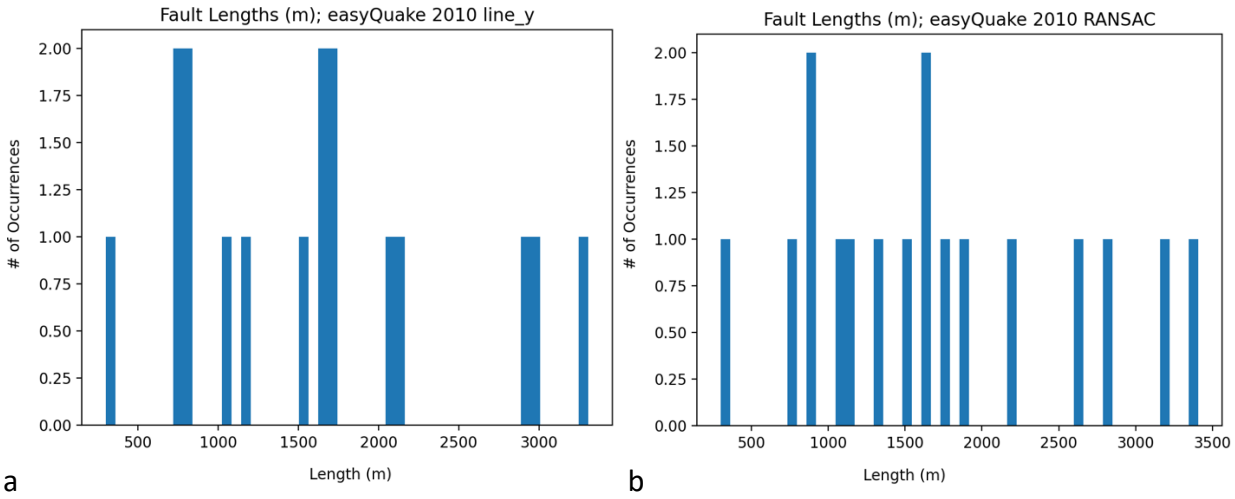


Figure 3.5.6 (top) Fault lineations for easyQuake 2010, relocated events.

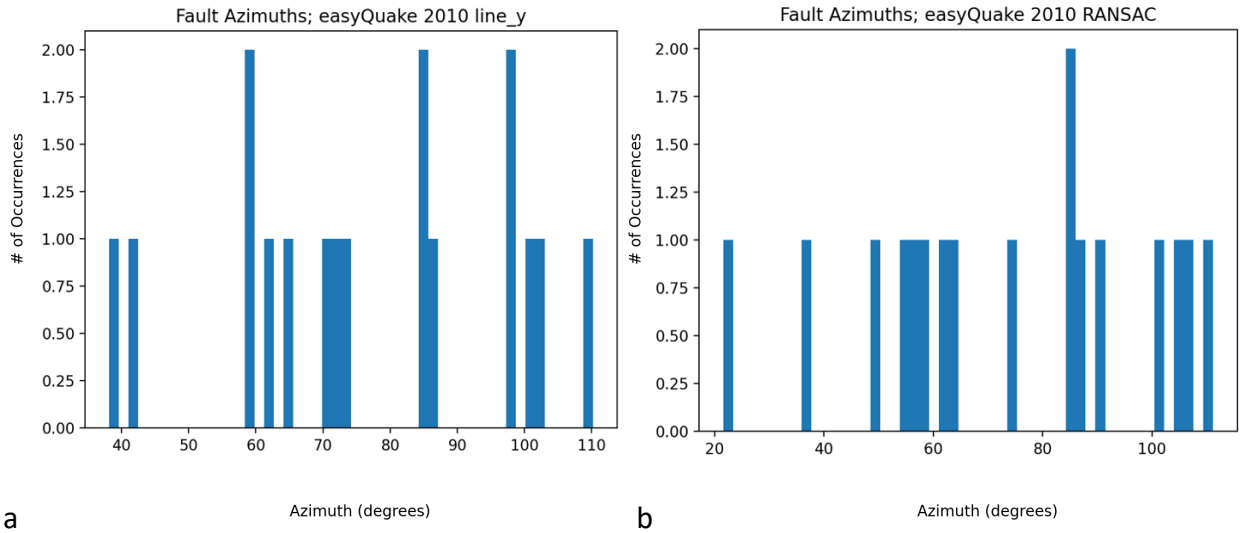
Figure 3.5.7 (bottom) Fault lineations for easyQuake 2010-2021, relocated events.

It was discussed in Chapter 2.5 that RANSAC was selected for creating the linear regressions as it is less susceptible to outliers without introducing more bias than a standard linear regression would. Figures 3.5.8 through 3.5.11 compare the standard linear regressions, "line_y," to RANSAC regressions.



a b

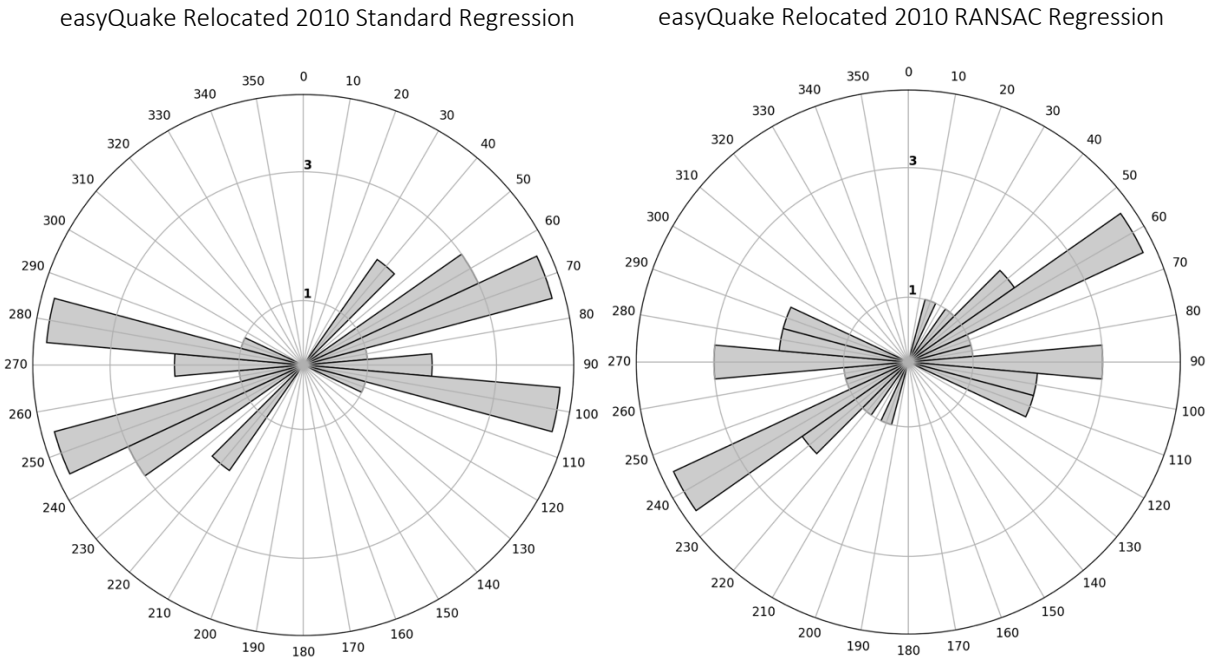
Figure 3.5.8 (a) Estimated fault lengths for easyQuake 2010 catalog based on standard linear regression (b) Estimated fault lengths for easyQuake 2010 catalog based on RANSAC



a b

Figure 3.5.9 (a) Estimated fault azimuths for easyQuake 2010 catalog based on standard linear regression (b) Estimated fault azimuths for easyQuake 2010 catalog based on RANSAC

Rose diagrams for the estimated fault orientations are shown below to represent the regional fault azimuth estimates more accurately, due to the bilateral nature of faults. Both rose diagrams show a general NE-SW trend for events. The standard regression shows orientations estimates to be predominantly N70°E and N100°E, while the RANSAC shows predominantly N60°E and N90°E.



Figures 3.5.10 and 3.5.11 Fig 3.5.9(left) is a rose diagram of estimated fault azimuths for the relocated easyQuake 2010 catalog based on standard linear regressions. Fig 3.5.10(right) is a rose diagram the same data but based on RANSAC regressions.

Since we are looking at faults based on earthquake occurrences, we know them to be seismogenic faults. A fault is most likely to be seismogenic if it is optimally oriented with relation to the regional stress field. Given this, the 2010 relocated easyQuake data's lineations indicate a regional stress field of predominantly N80°E.

Both diagrams' results have a general agreement with accepted S_{Hmax} measurements, but the RANSAC results are more closely aligned with the N68°E and N59°E range discussed in

Skoumal et al., 2019 than are the standard regression results. When the data frame is expanded to include 2010 through 2021 data though, we see the regional stress field is indicated to more predominantly be aligned with N80°E.

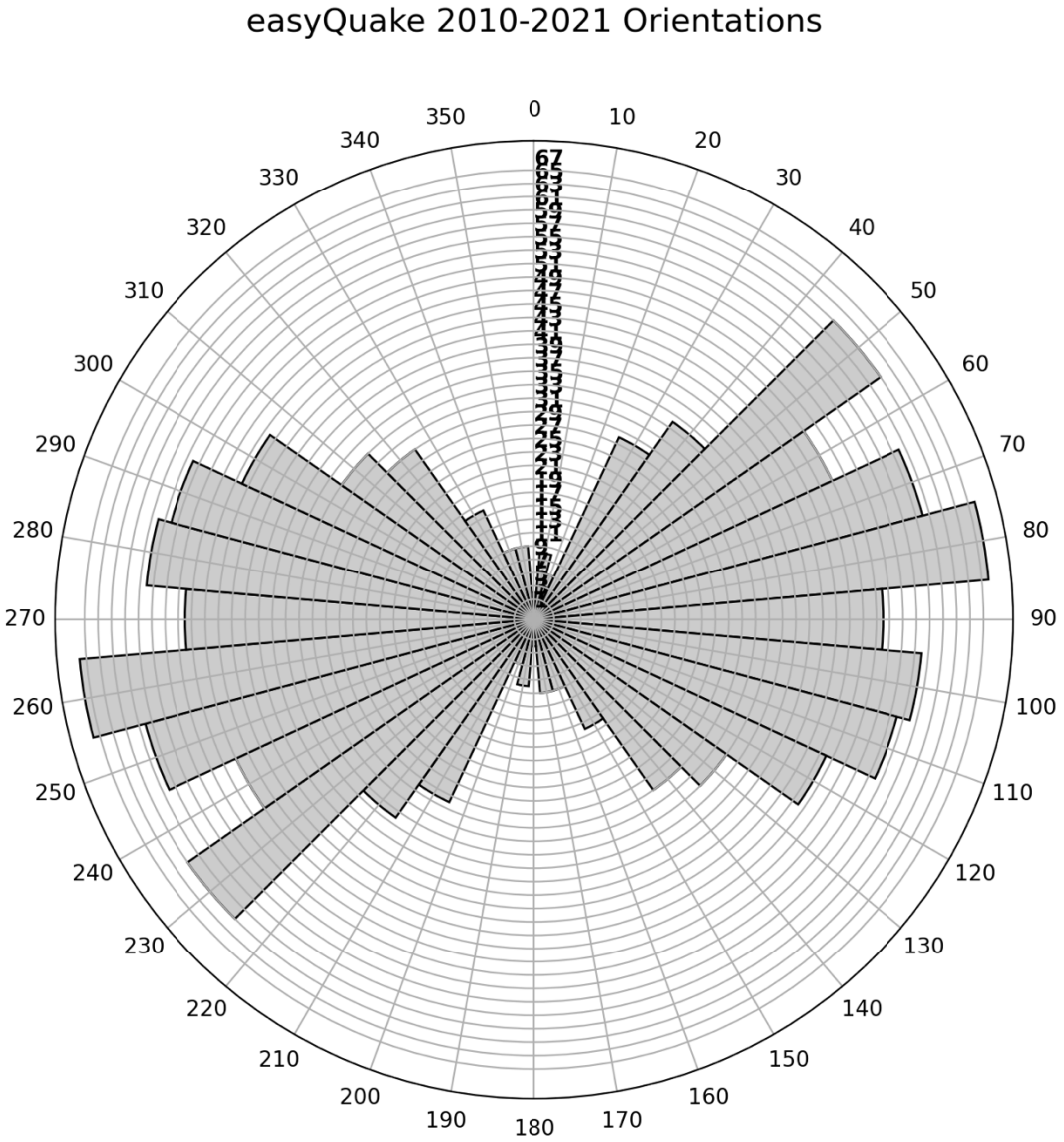


Figure 3.5.12 Rose diagram for the relocated easyQuake catalog for 2010 through 2021.

3.6 Examining a Broad Region

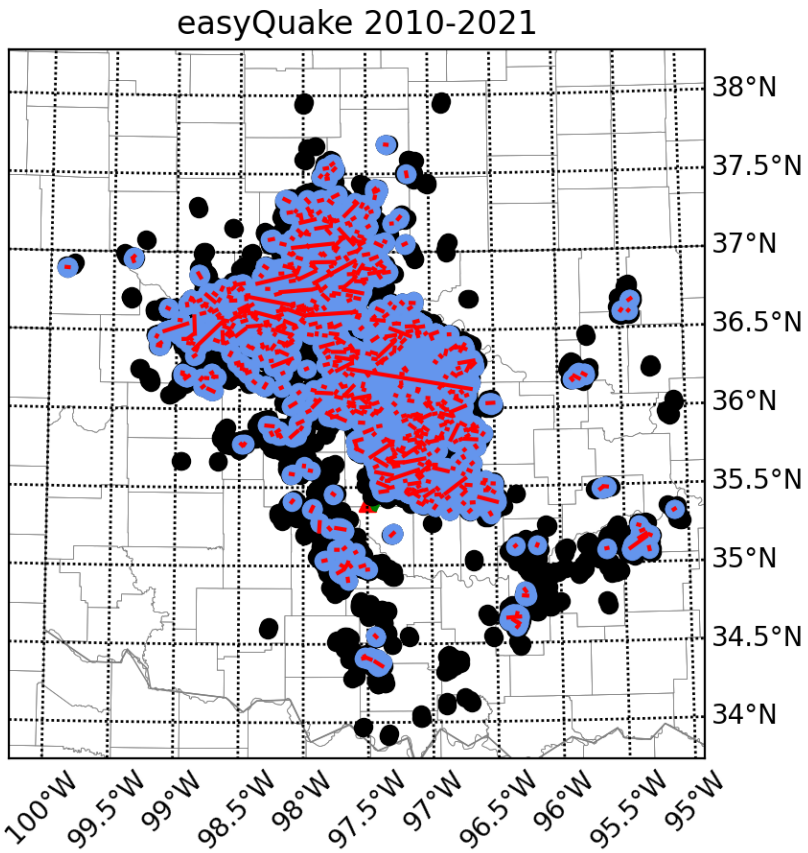


Figure 3.6.1 Delineated clusters for the relocated 2010-2021 easyQuake catalog. Blue dots represent events belonging to clusters. Black dots represent noise point events. Red lines show cluster regression lines.

The expanded selection of relocated easyQuake data is substantially larger than the dataset limited to the Jones area in only 2010. We examine this larger dataset to look at the broader region, rather than local to one city one year. This helps examine seismic activity for the state as a whole over time.

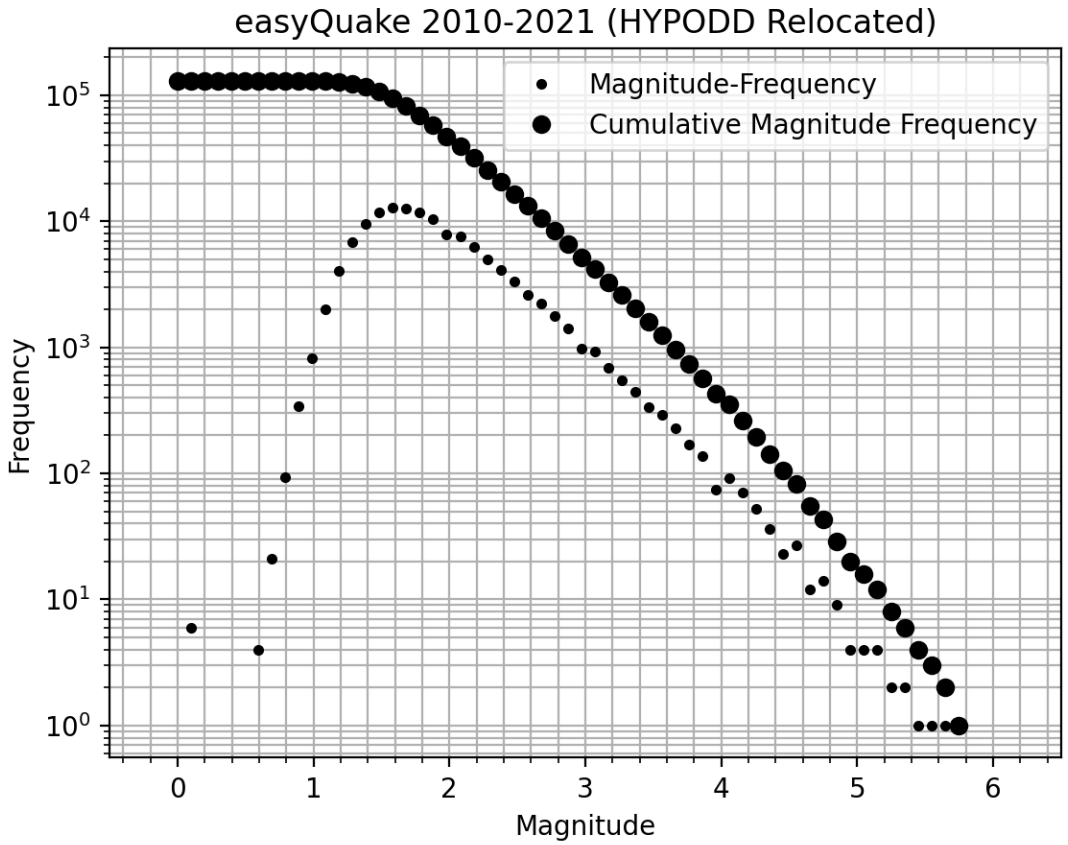


Figure 3.6.2 Gutenberg Richter plot for the 2010 through 2021 easyQuake data.

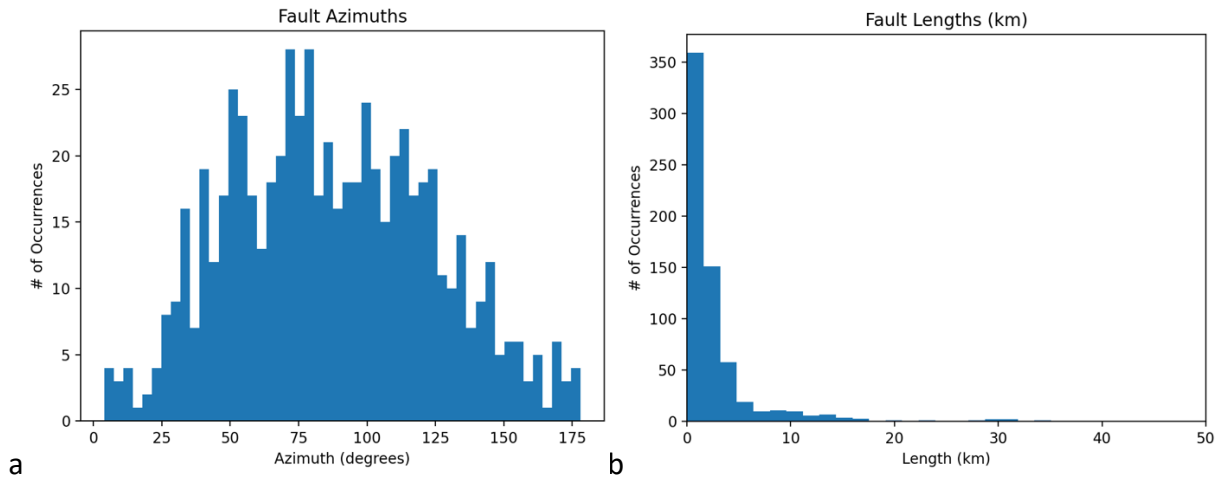


Figure 3.6.2 (a) Estimated fault azimuths for the relocated 2010 through 2021 easyQuake data (b) Estimated fault lengths for the relocated 2010 through 2021 easyQuake data. (b) shows the relationship of fault length to occurrence is roughly logarithmic, just as is the relationship between event magnitudes and occurrence rates.

3.7 Propagation Rates

To examine the concept of event propagation along faults, we look at the locations of events relative to their corresponding fault. If propagation is indeed occurring, we would expect to see events located farther along the fault as time goes on. In Figure 3.6.2, we see several faults with repeated events at the same locations over time. There are also several instances of faults with seismic activity along the entire length at once. Due to the compound estimates in this workflow, we see some fault lengths longer than would be expected, as seen in Figure 3.7.1.

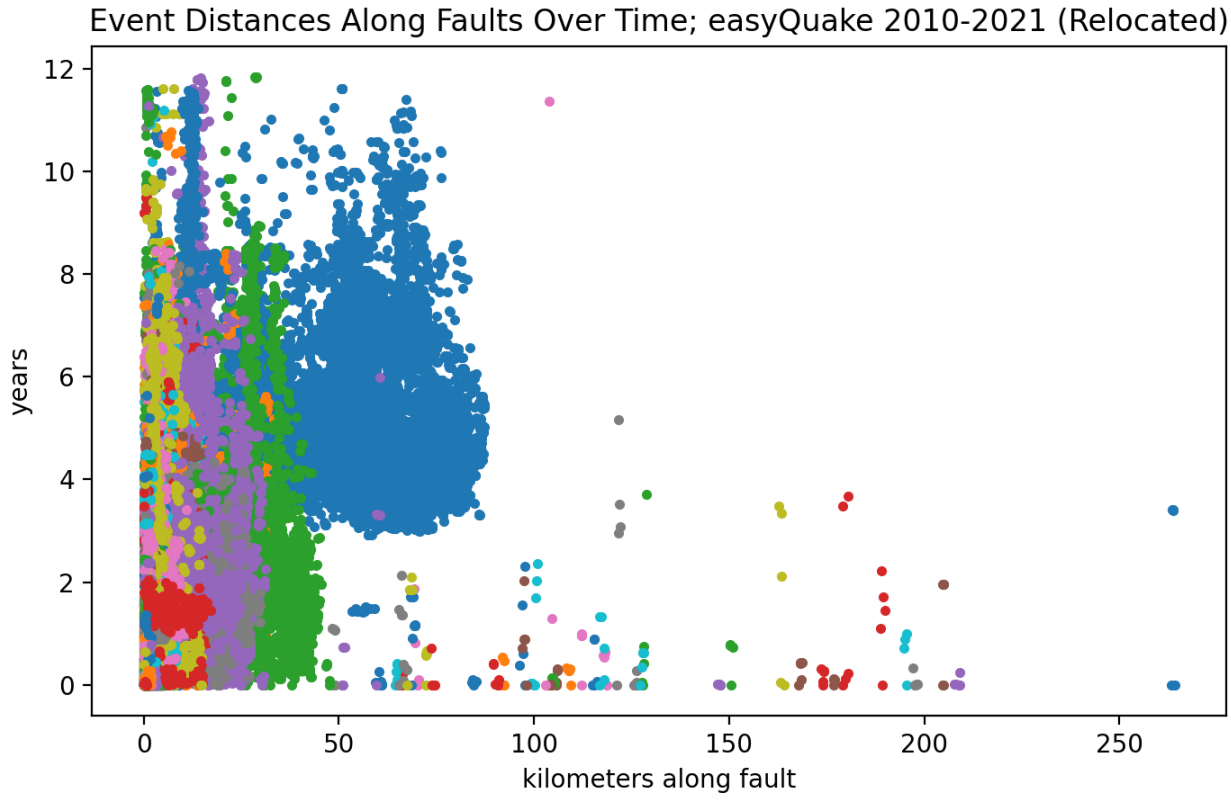


Figure 3.7.1 Distance along faults for events over a period of twelve years, 2010 through 2021. This plot displays relocated easyQuake data. Events belonging to the same fault are assigned the same color.

These figures do show fault lengths much longer than we would expect. These are generally artifacts of the clustering. Since there are several steps of parameter selections and adjustments, there is a compoundment of estimations, which can lead to a few outlying results such as this. It is difficult for DBSCAN to properly cluster datasets this large, which is why it is particularly helpful to examine individual years at a time as at the beginning of this study.

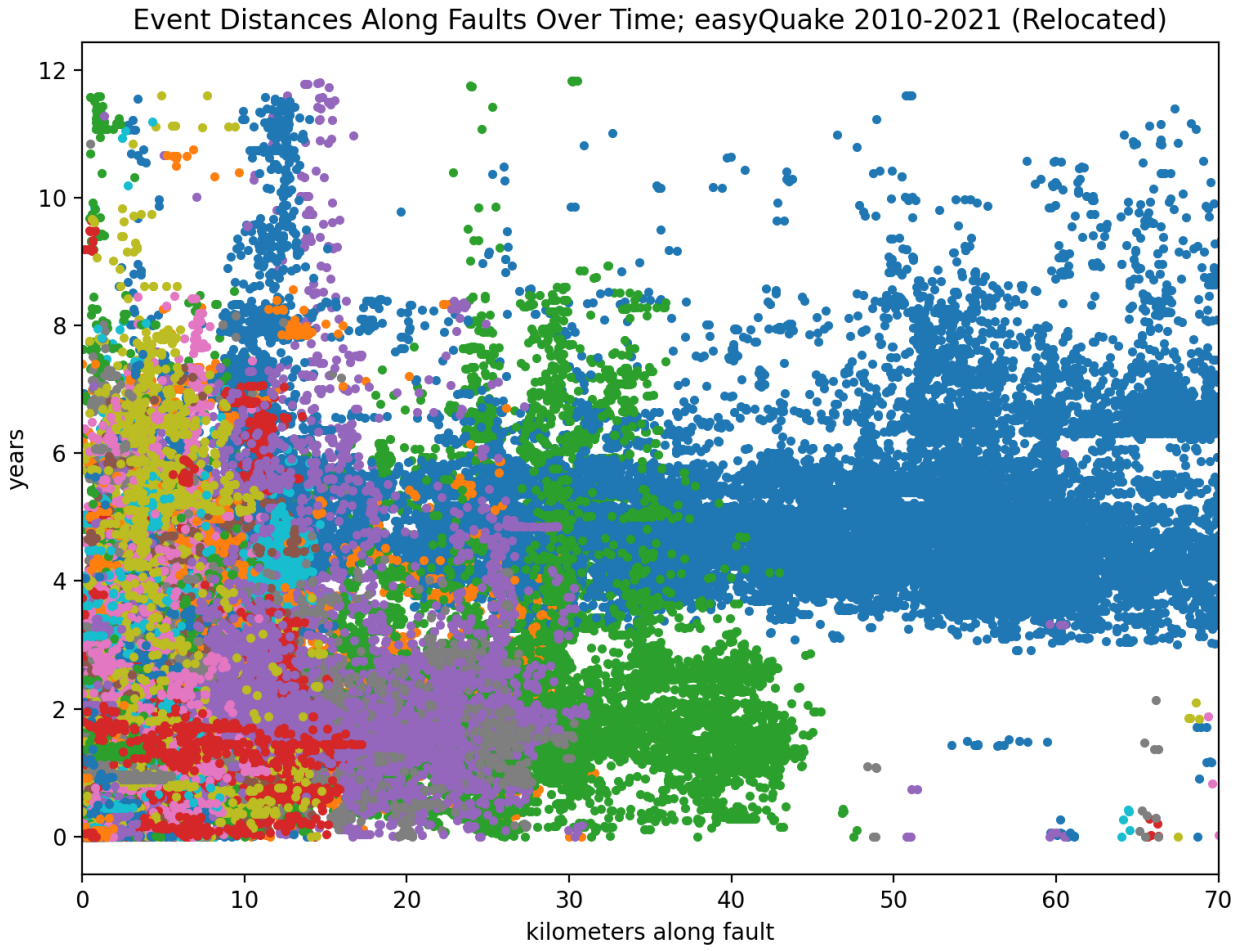


Figure 3.7.2 Distance along faults for events over a period of twelve years, 2010 through 2021. This plot displays relocated easyQuake data and is limited to 70km fault distances and smaller. Events belonging to the same fault are assigned the same color. We can see instances of events at multiple distances along fault lengths at the same time which would indicate that more of the fault is rupturing at once than just individual locations.

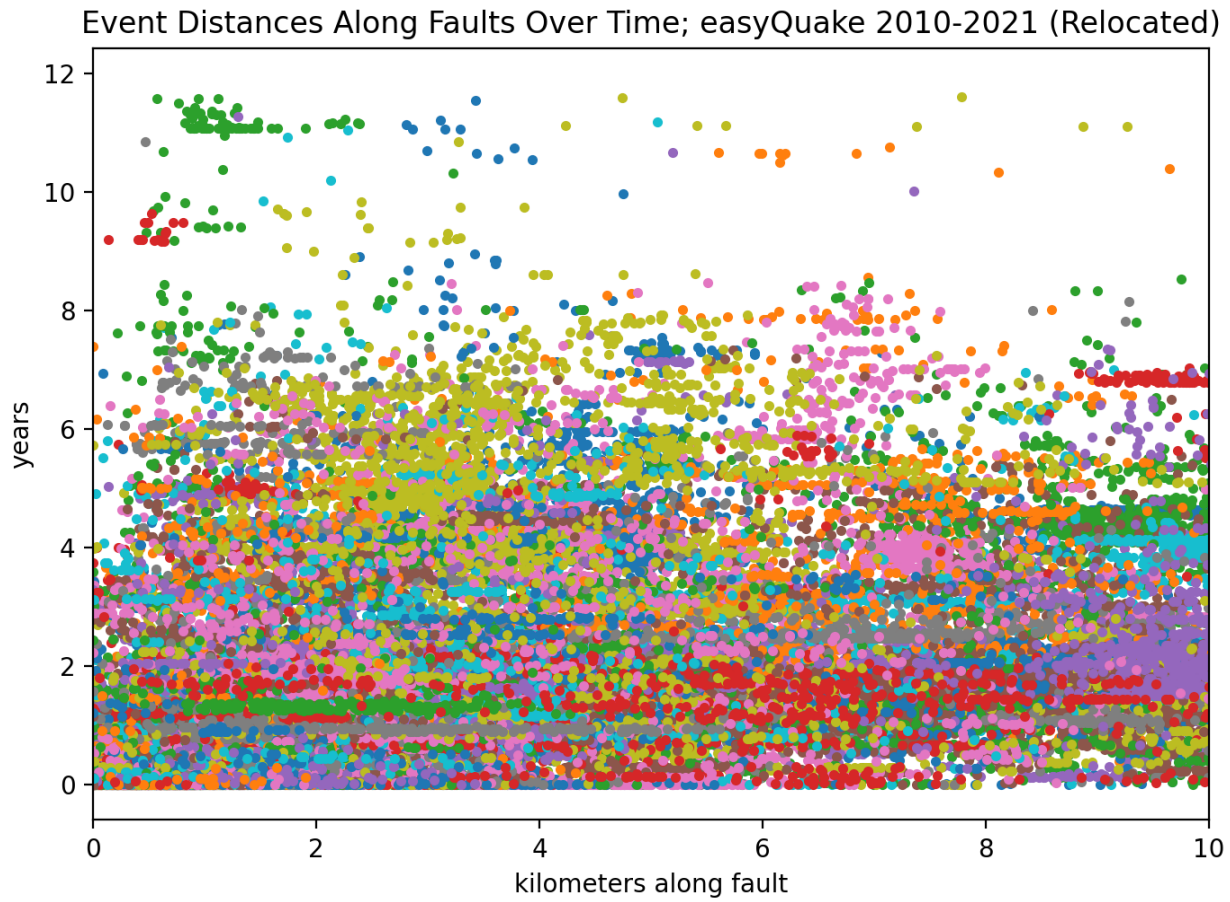


Figure 3.7.3 Distance along faults for events over a period of twelve years, 2010 through 2021. This plot displays relocated easyQuake data and is limited to 10km fault distances and smaller. Events belonging to the same fault are assigned the same color. This figure shows that over time, events are occurring at different locations along their respective faults, but that there is no clear migration pattern at this time.

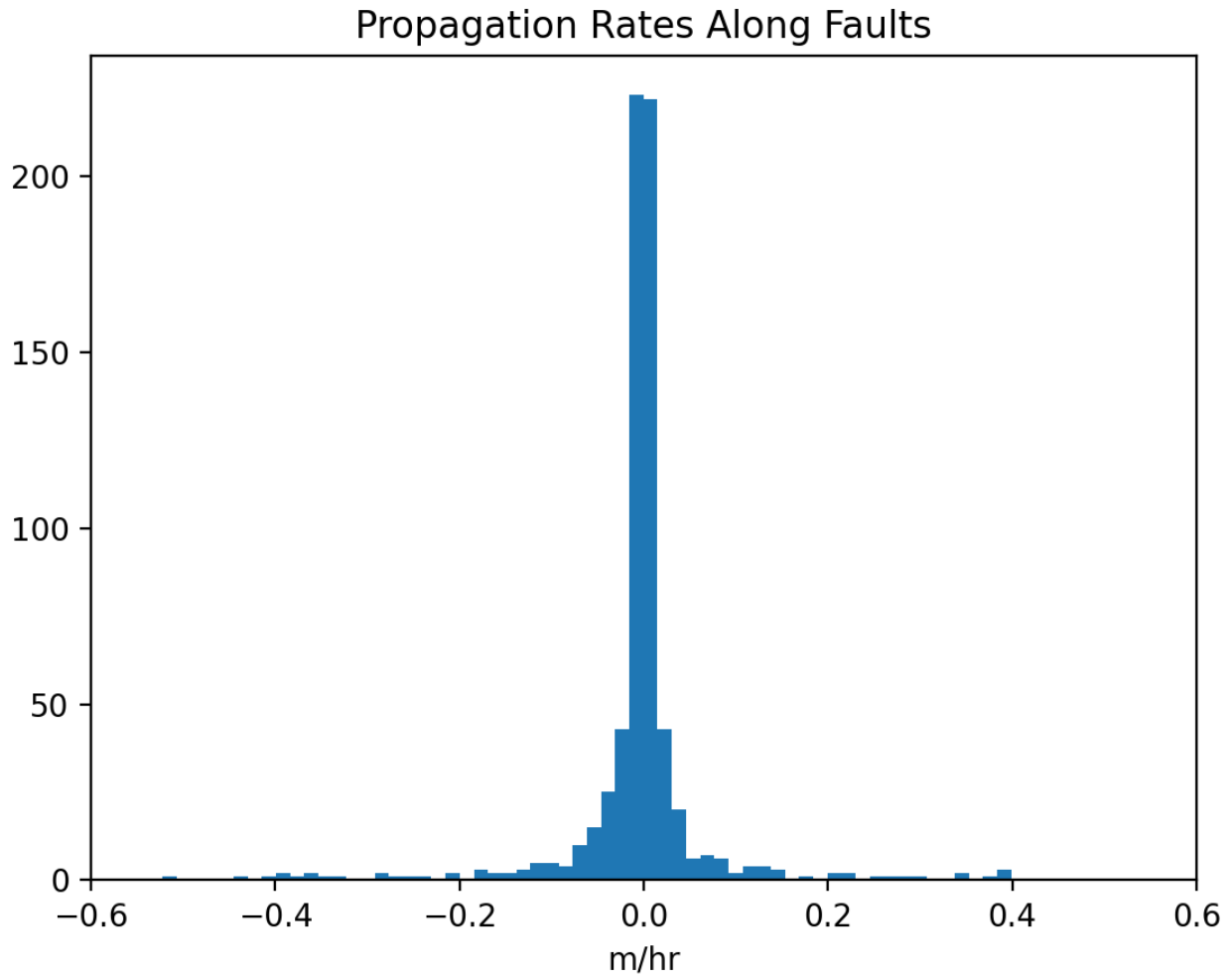


Figure 3.7.4 Histogram of event propagation rates along faults in meters per hour. Event occurrences are migrating along their respective fault segments at a rate most typically averaging near 0.0 meters per hour.

Figure 3.7.4 shows the propagation rates of events along faults found with the relocated easyQuake data for 2010 through 2021. It has a roughly symmetrical distribution centered right around 0.0 meters per hour. A “negative” propagation rate indicates that events are occurring closer the earliest event on the fault, not that the propagation is slowing down. The propagation rate is with respect to the earliest event on a fault; Westward movement is denoted as negative, while Eastward movement is denoted as positive. Over 200 of the faults

found have propagation rates in this range. Overall event propagation along faults in Oklahoma for the last twelve years would be slow, as events tend to stay centered around the earliest event on a fault based on the histogram in Figure 3.7.4.

CHAPTER 4: DISCUSSION AND CONCLUSIONS

Overview

- Machine learning techniques for earthquake catalog analysis are continuously developing.
- A more complete catalog yields more information about a region's seismic activity, which can detect and reveal patterns that would otherwise be missed.
- An automated system for identifying faults and their orientations can help improve the seismic hazard map more quickly and efficiently than manual methods.
- Historical data can be useful for current regional studies as advancements in machine learning techniques have expanded the information it is possible to glean from the archived data.

It is important to study seismogenic faults and their geometries to better understand the potential for fault slippage and therefore, potential seismic activity. A strike-slip fault that is optimally oriented with respect to the principal tectonic stress direction of the region has an increased susceptibility to slipping. If an area has a large number of these optimally oriented faults, there is a higher potential seismic hazard, so it is beneficial to have as complete of a fault map as possible to help mitigate risk. This workflow begins with earthquake catalog data, so we know the faults identified are seismogenic. This means that faults identified with this method are more likely to be optimally oriented faults and have a higher level of seismic hazard. Orientations of faults that are visible at the surface can be determined via field mapping, such

as the exposed faults and fracture systems mapped in Kolawole et al., 2019, as well as manually determined from subsurface images, as is also done in Kolawole et al., 2019. These types of methods can be time consuming and more susceptible to human error than an automated identification system.

Skoumal et al., 2019 utilized a system they have dubbed the FaultID algorithm. FaultID is what this study has been modelled after to identify potential fault lineations and improve the completeness and accuracy of the regional fault map. As was found in the Skoumal et al., (2019) study, this system delineates potential fault segments that were previously unmapped. This is done without field mapping or subsurface imaging. The resulting segments in the Skoumal19 study are generally aligned with the regional stress field, as are the segments identified in this study, which means these faults are what would be considered “optimally oriented”. However, that study utilized a regional network based matched-filter catalog which would be biased towards only detecting events nearby existing events. Applying the general methodology of clustering and identifying lineations with a machine-learning derived catalog eliminates that bias.

Our methodology can be utilized with any earthquake catalog available and can be expanded to explore other regions of interest. It works to create an automated workflow for mapping seismogenic faults. Resulting fault maps can then be used for other studies and analyzing regional characteristics such as stress state, as in Qin et al., 2019. Adjustments need to be made to each parameter selection based on the region’s geology; for example, HYPODD will need an updated velocity model input to reflect the region’s subsurface, as well as clustering and regression parameter adjustments based on available data.

Oklahoma has more microseismicity than are typically recorded by the current seismic monitoring system as evidenced in Figure 1.1.1. Machine learning techniques, such as those implemented with easyQuake, are improving the data collection and analysis process, as it captures more events and corresponding spatio-temporal data than the USGS network alone. Events with small magnitudes – ≤M2.0 – are often not captured by the current networks and USGS does not report events with magnitudes below M2.5. Figures 3.1.1 through 3.1.6 are Gutenberg-Richter plots of these catalogs, which show easyQuake as having the smallest magnitude of completeness of the available datasets. Data coverage is generally lacking for events smaller than M0, but advancements in techniques such as easyQuake can help lessen these deficiencies.

Other regional studies show events which were potentially induced by high-volume wastewater injection, as well as possible migrations due to the increased pore pressure from these injections. As in Keranen et al., 2014, catalog data is used to plot the distances of events to Southeast Oklahoma City disposal wells. These distances are used to monitor fault segment orientation and potential event sequence migrations with relation to a high-volume disposal well. In Figures 3.3.1 and 3.3.3 there is a ramp-like feature in the dataset that would indicate a migration of events away from the well over time, as Keranen suggests. The same structure is seen in all four of the analyzed catalogs. The feature is less apparent in the easyQuake data, however, than in the Keranen14 data. When expanded to examine distances beyond 50km from the well, the ramp structure is no longer prominent. Keranen14 asserts that the high-volume wastewater injection is increasing the pore pressure in the area and inducing seismicity as the pore pressure field diffuses across the fault damage zone. While this assertion seems

reasonable, more research is needed to verify the other assertion that the injections from on one specific site are causing event migrations along the faults suggested in Keranen et al., 2014. The Keranen14 study suggests a propagation rate of approximately ten kilometers from the beginning of June to the beginning of August 2010, based on Figure 1.1.2c. This is equivalent to 1 kilometer per 6.1 days, or ~0.007 km/hr. Given the slow rates shown in Figure 3.7.4, this does not seem to be an unreasonable rate.

Several faults seem to have ruptured along their entire lengths on multiple occasions, as seen in Figures 3.7.1 through 3.7.3. Whether specific events caused these ruptures could be studied in the future and may be made easier to determine with the implementation of a system for rapid automated fault identification such as tested in this study.

Hydraulic loading pressure has been cited as a source of induced seismicity in several regions of the world, such as for reservoir impoundment (e.g., Costain & Bollinger, 2010, Foulger et al., 2017, Grasso et al., 2018). This hypothetical cause was tested for regional reservoirs versus seismicity to determine the possibility of the increased seismicity being due to local reservoir levels. There is a clear lack of correlation between nearby reservoir levels with local seismic rates. The reservoirs' levels' movements mimic each other, but seismicity rates in this area do not follow the same trend. The seismicity for the area is fairly consistent, with a sharp increase between 2013 and 2014. The beginning of said sharp increase coincides with the spike in both reservoirs' gage heights, but as the gage heights decrease, the seismicity rate continues to increase. Based on this data, these reservoirs do not appear to be inducing seismicity in the Jones area, as the reservoir gage heights and local seismicity levels would be expected to have a much higher degree of correlation if this were the case.

After successfully testing this system on the 2010 subset of Oklahoma catalog data, we expanded the study area to the entire state of Oklahoma. This expansion results in over 700 potential fault segments, as shown in Figure 3.5.6, and represented again in Figure 3.6.1. These segments were able to be identified in a short amount of time, with no reliance on surface mapping or subsurface imaging. The average fault length is 2.87km. The largest fault we identified through our clustering is approximately 87km. Future studies could implement a workflow such as the one we proposed and extend it by including a relative stress inversion (e.g., Qin et al., 2019) for seismic hazard studies in a specific region. These types of studies, coupled with other passive-source methodologies such as shear wave splitting (e.g., Ortega-Romo et al., 2021), could be undertaken at the small to regional scale prior to the start of unconventional hydrocarbon activities in order to couple the delineation of fault geometries with inferences of the regional stress field.

BIBLIOGRAPHY

1. "ANSS - Advanced National Seismic System." *ANSS - Advanced National Seismic System |U.S. Geological Survey*, https://www.usgs.gov/programs/earthquake-hazards/anss-advanced-national-seismic-system

2. Araya, María C.. (2017). Application of the double difference earthquake relocation algorithm methodology using hypoDD at four seismic sequences in Costa Rica. Revista Geológica de América Central, (57), 7-21. https://dx.doi.org/10.15517/rgac.v0i57.30041

3. Barbour, Andrew J., et al. "The Effects of Varying Injection Rates in Osage County, Oklahoma, on the 2016*m*w 5.8 Pawnee Earthquake." *Seismological Research Letters*, vol. 88, no. 4, 2017, pp. 1040–1053., https://doi.org/10.1785/0220170003 https://scits.stanford.edu/sites/g/files/sbiybj22081/files/media/file/0220170003.full__0.pdf

4. Chen, C., and A. A. Holland, (2016). PhasePApy: A Robust Pure Python Package for Automatic Identification of Seismic Phases, Seismological Research Letters, 87(6), DOI: 10.1785/0220160019 https://doi.org/10.1785/0220160019

5. Chen, X., N. Nakata, C. Pennington, J. Haffener, J. Chang, X. He, Z. Zhan, S. Ni, and J. Walter (2017). The Pawnee earthquake as a result of the interplay among injection, faults and foreshocks, Sci. Rep. 7, no. 1, 4945. https://www.nature.com/articles/s41598-017-04992-z

6. Chen, X., Shearer, P., and Abercrombie (2012). Spatial migration of earthquakes within seismic clusters in Southern California: Evidence for fluid diffusion. JGR Solid Earth Vol 117, no. B4, 2012. https://doi.org/10.1029/2011JB008973

7. Collins, Robert. *Robust Estimation: RANSAC CSE486, Penn State*. https://www.cse.psu.edu/~rtc12/CSE486/lecture15.pdf

8. Costain, J. K., and G. A. Bollinger. "Review: Research Results in Hydroseismicity from 1987 to 2009." *Bulletin of the Seismological Society of America*, vol. 100, no. 5A, 2010, pp. 1841–1858., https://doi.org/10.1785/0120090288 https://www.researchgate.net/publication/228349499_Review_Research_Results_in_Hydroseismicity_from_1987_to_2009

9. Daniel T. Trugman and Peter M. Shearer (2017) GrowClust: A Hierarchical Clustering Algorithm for Relative Earthquake Relocation, with Application to the Spanish Springs and Sheldon, Nevada, Earthquake Sequences. Seismological Res. Lett. 88 - 2A, 379-391, 2017. doi:10.1785/0220160188 https://igppweb.ucsd.edu/~shearer/mahi/PDF/2017/Trugman_growclust_2017.pdf

10. DBSCAN Documentation, "Sklearn.cluster.DBSCAN." *Scikit*, https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html

11. Dilts, Ben, and Karl Sun. "Intelligent Diagramming." *Lucidchart*, https://www.lucidchart.com/pages/

12. Deichmann, N. and Garcia-Fernandez, M. (1992) Rupture geometry from high-precision relative hypocentre locations of microearthquake clusters. Geophys. J. Inr. (1992) 110, 501-517 https://www.academia.edu/48972750/Rupture_geometry_from_high_precision_relative_hypocentre_locations_of_microearthquake_clusters

13. Dudik, Joshua M et al. "A comparative analysis of DBSCAN, K-means, and quadratic variation algorithms for automatic identification of swallows from swallowing accelerometry signals." *Computers in biology and medicine* vol. 59 (2015): 10-18. doi: 10.1016/j.compbiomed.2015.01.007 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4363248/

14. Dunn, Meredith 2004. Relocation of Eastern Tennessee Earthquakes Using hypoDD. Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of Master of Science in Geophysics in the Department of Geosciences. https://vtechworks.lib.vt.edu/bitstream/handle/10919/10085/MDunnthesis.pdf?sequence=1&isAllowed=y

15. Ester, Martin, et al. 1996 *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise* https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf

16. Fischler, Martin A., and Robert C. Bolles. "Random Sample Consensus." *Communications of the ACM*, vol. 24, no. 6, 1981, pp. 381–395., https://doi.org/10.1145/358669.358692

17. Foulger, Gillian R., et al. "Global Review of Human-Induced Earthquakes." *Earth-Science Reviews*, Elsevier, 26 July 2017, https://www.sciencedirect.com/science/article/pii/S001282521730003X?via%3Dihub

18. Grasso, J.-R., et al. "Patterns of Reservoir-Triggered Seismicity in a Low-Seismicity Region of France." *Bulletin of the Seismological Society of America*, vol. 108, no. 5B, 2018, pp. 2967–2982., https://doi.org/10.1785/0120180172 https://www.researchgate.net/publication/327877345_Patterns_of_Reservoir-Triggered_Seismicity_in_a_Low-Seismicity_Region_of_France

19. Hardebeck, Jeanne L., Shearer, Peter M.; A New Method for Determining First-Motion Focal Mechanisms. *Bulletin of the Seismological Society of America* 2002; 92 (6): 2264–2276. doi: https://doi.org/10.1785/0120010200

20. Holland, Austin. “Austinholland/Phasepapy: Python Seismic Phase Picker and Associator.” *GitHub*, https://github.com/austinholland/PhasePApy.

21. Hutton, L. K. and Boore, David M. (1987) *The M_L scale in Southern California.* Bulletin of the Seismological Society of America, 77 (6). pp. 2074-2094. ISSN 0037-1106. https://resolver.caltech.edu/CaltechAUTHORS:20140905-113510505

22. “Hypoinverse Earthquake Location.” *HYPOINVERSE Earthquake Location | U.S. Geological Survey*, https://www.usgs.gov/software/hypoinverse-earthquake-location

23. Jacob I. Walter, Paul Ogwari, Andrew Thiel, Fernando Ferrer, Isaac Woelfel; easyQuake: Putting Machine Learning to Work for Your Regional Seismic Network or Local Earthquake Study. *Seismological Research Letters* 2020;; 92 (1): 555–563. doi: https://doi.org/10.1785/0220200226

24. Kanamori, Hiroo, and Emily E Brodsky. “The Physics of Earthquakes.” *Reports on Progress in Physics*, vol. 67, no. 8, 2004, pp. 1429–1496., https://doi.org/10.1088/0034-4885/67/8/r03

25. Kazemi-Beydokhti, M., Ali Abbaspour, R. & Mojarab, M. Spatio-Temporal Modeling of Seismic Provinces of Iran Using DBSCAN Algorithm. *Pure Appl. Geophys.* 174, 1937–1952 (2017). https://doi.org/10.1007/s00024-017-1507-0

26. Keranen, K. M., Weingarten, M., Abers, G. A., Bekins, B. A., and Ge, S. (2014) Sharp increase in central Oklahoma seismicity since 2008 induced by massive wastewater injection. Science 345 (6195), 448-451. DOI: 10.1126/science. 1255802 originally published online July 3, 2014 https://wpg.forestry.oregonstate.edu/sites/wpg/files/seminars/keranen_etal_2014_induced_seismicity.pdf

27. Klein, Fred W. “User's Guide to Hypoinverse-2000, a FORTRAN Program to Solve for ... - USGS.” *USGS.gov*, UNITED STATES DEPARTMENT OF THE INTERIOR GEOLOGICAL SURVEY, June 2014, https://escweb.wr.usgs.gov/content/software/HYPOINVERSE/doc/hyp1.40.pdf

28. Krischer, Lion, et al. “ObsPy: A Bridge for Seismology into the Scientific Python Ecosystem.” *Computational Science & Discovery*, vol. 8, no. 1, 2015, p. 014003., https://doi.org/10.1088/1749-4699/8/1/014003

29. Kolawole, F., et al. “The Susceptibility of Oklahoma’s Basement to Seismic Reactivation.” *Nature Geoscience*, vol. 12, no. 10, 2019, pp. 839–844., https://doi.org/10.1038/s41561-019-0440-5 https://www.nature.com/articles/s41561-019-0440-5.pdf

30. Lawson, J. E., and K. V. Luza (1995). Earthquake map of Oklahoma (with explanatory text), Oklahoma Geol. Surv. Geologic Map 35, scale 1:500,000. http://ogs.ou.edu/docs/geologicmaps/GM35M.pdf

31. Mesimeri, Maria, Karakostas, Vassilios, Papadimitriou, Eleftheria, Tsaklidis, George, and Jacobs, Katrina. Relocation of recent seismicity and seismotectonic properties in the Gulf of Corinth (Greece). Geophysical Journal International, Oxford University Press (OUP), 2018, 212 (2), pp.1123-1142. ff10.1093/gji/ggx450ff. ffhal-01676804f https://academic.oup.com/gji/article/212/2/1123/4554391?login=true

32. McNamara, D. E., H. M. Benz, R. B. Herrmann, E. A. Bergman, P. Earle, A. Holland, R. Baldwin, and A. Gassner (2015), Earthquake hypocenters and focal mechanisms in central Oklahoma reveal a complex system of reactivated subsurface strike-slip faulting, Geophys.Res. Lett.,42, 2742–2749, doi:10.1002/2014GL062730 https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2014GL062730

33. McNamara, D. E., Rubinstein, J. L., et al. "Efforts to Monitor and Characterize the Recent Increasing Seismicity in Central Oklahoma."(2015) *The Leading Edge*, https://library.seg.org/doi/10.1190/tle34060628.1

34. Mousavi, S.M., Ellsworth, W.L., Zhu, W. *et al.* Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nat Commun* 11, 3952 (2020). https://doi.org/10.1038/s41467-020-17591-w

35. *NumPy*, https://numpy.org/

36. "ObsPy Documentation (1.3.0)Á." *ObsPy Documentation (1.3.0) - ObsPy 1.3.0 Documentation*, 10 Mar. 2022, https://docs.obspy.org/

37. ObsPy. "ObsPy: A Python Toolbox for Seismology/Seismological Observatories." *GitHub*, https://github.com/obspy/obspy

38. Ortega-Romo, A. D., J. I. Walter, X. Chen, and B. M. Carpenter (2021), Spatially distinct tectonic zones across Oklahoma inferred from shear wave splitting, Seismol. Res. Lett., doi:10.1785/0220200237 https://www.researchgate.net/publication/351692754_Spatially_Distinct_Tectonic_Zones_across_Oklahoma_Inferred_from_Shear-Wave_Splitting

39. Ross, Z.E.; Meier, M.; Hauksson, E.; Heaton, T.H.; Generalized Seismic Phase Detection with Deep Learning. *Bulletin of the Seismological Society of America* 2018; 108 (5A): 2894–2901. doi: https://doi.org/10.1785/0120180080 https://arxiv.org/pdf/1805.01075.pdf

40. Rosson, Z, et al. *Narrow Spatial Aftershock Zones for Induced Earthquake Sequences in Oklahoma* Geophysical Research Letters, https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019GL083562

41. Schoenball, M., & Ellsworth, WL. (2017). Waveform-relocated earthquake catalog for Oklahoma and southern Kansas illuminates the regional fault network. *Seismological Research Letters*, 88(5), 1252-1258. http://dx.doi.org/10.1785/0220170083

42. Schoenball, M., Walsh, FR, Weingarten, M., & Ellsworth, WL. (2018). How faults wake up: The Guthrie-Langston, Oklahoma earthquakes. *Leading Edge*, 37(2), 100-106. http://dx.doi.org/10.1190/tle37020100.1 Retrieved from https://escholarship.org/uc/item/09n7q9pt

43. Pedregosa *et al.*, (2011). Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830, 2011.

44. Skoumal, R. J., M. R. Brudzinski, and B. S. Currie (2015). Microseismicity induced by deep wastewater injection in southern Trumbull County, Ohio, Seismol. Res. Lett. 86, 1326–1334, doi: 10.1785/0220150055 https://www.researchgate.net/publication/282965959_Microseismicity_Induced_by_Deep_Wastewater_Injection_in_Southern_Trumbull_County_Ohio

45. Skoumal, R., Kaven, J., Walter, J. (2019) Characterizing Seismogenic Fault Structures in Oklahoma Using a Relocated Template-Matched Catalog. *Seismological Research Letters,* (2019) 90 (4): 1535–1543. https://doi.org/10.1785/0220190045

46. Spyder Team. "Home - Spyder Ide." *Home - Spyder IDE*, https://www.spyder-ide.org/

47. Waldhauser, F., HypoDD: A computer program to compute double-difference earthquake locations, U.S. Geol. Surv. Open-File Report, 01-113, Menlo Park, California, 2001. https://www.ldeo.columbia.edu/~felixw/papers/Waldhauser_OFR2001.pdf

48. Waldhauser, F. and Ellsworth, W. (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California, Bull. Seismol. Soc. Am. 90, no. 6, 1353–1368, doi: 10.1785/0120000006 https://www.ldeo.columbia.edu/~felixw/papers/Waldhauser_Ellsworth_BSSA2000.pdf

49. Waldhauser, F., and D. P. Schaff (2008). Large-scale relocation of two decades of northern California seismicity using cross-correlation and double-difference methods, J. Geophys. Res. 113, no. B8, doi: 10.1029/2007JB005479 https://www.ldeo.columbia.edu/~felixw/papers/Waldhauser_Schaff_JGR2008.pdf

50. Walter, Jacob I., et al. "The Oklahoma Geological Survey Statewide Seismic Network." *Seismological Research Letters*, Seismological Society of America, 13 Nov. 2019, https://www.osti.gov/pages/servlets/purl/1574703

51. Walter, J. I., P. Ogwari, A. Thiel, F. Ferrer, and I. Woelfel (2020). easyQuake: Putting Machine Learning to Work for Your Regional Seismic Network or Local Earthquake Study, Seismol. Res. Lett. XX, 1–9, doi: 10.1785/0220200226 https://www.jakewalter.net/papers/Walteretal2020_easyquake.pdf

52. Walter, J. I., P. Ogwari, A. Thiel, F. Ferrer, and I. Woelfel (2021), easyQuake: Putting machine learning to work for your regional seismic network or local earthquake study, Seismological Research Letters, 92(1): 555–563, https://doi.org/10.1785/0220200226

53. Wang, Y., G. Ouillon, J. Woessner, D. Sornette, and S. Husen (2013). Automatic reconstruction of fault networks from seismicity catalogs including location uncertainty, J. Geophys. Res. 118, no. 11, 5956–5975 https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2013JB010164

54. Wells, Donald L, and Kevin J Coppersmith. *New Empirical Relationships among Magnitude, Rupture Length, Rupture ...* Bulletin of the Seismological Society of America, Vol. 84, No. 4, Pp. 974-1002, Aug. 1994, https://www.resolutionmineeis.us/sites/default/files/references/wells-coppersmith-1994.pdf.

55. Qin, Y., Chen, X., Walter, J. I., Haffener, J., Trugman, D. T., Carpenter, B. M., et al. (2019). Deciphering the stress state of seismogenic faults in Oklahoma and southern Kansas based on an improved stress map. Journal Geophysical Research: Solid Earth,124, 12,920-12,934 https://doi.org/10.1029/2019JB018377

56. Xu, Rui, and Donald Wunsch. "Survey of Clustering Algorithms." *IEEE Xplore*, IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 16, NO. 3, MAY 2005 645, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1427769

APPENDIX

For python scripts and HYPODD input:
https://github.com/kaycee-sims