UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ROLES OF FLUID INJECTION, STRESS STATE, GEOLOGICAL STRUCTURE, AND EARTHQUAKE INTERACTION ON OKLAHOMA EARTHQUAKES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

YAN QIN
Norman, Oklahoma
2020

ROLES OF FLUID INJECTION, STRESS STATE, GEOLOGICAL STRUCTURE, AND EARTHQUAKE INTERACTION ON OKLAHOMA EARTHQUAKES

A DISSERTATION APPROVED FOR THE
SCHOOL OF GEOSCIENCES

BY THE COMMITTEE CONSISTING OF

Dr. Xiaowei Chen, Chair

Dr. Brett Carpenter

Dr. Michael Behm

Dr. Jacob Walter

Dr. Xingru Wu

TABLE OF CONTENTS

Table of Contents iv

List of Figures vii

List of Tables xvi

Acknowledgements xvii

Vita and Publications xviii

Abstract of the Dissertation xix

1 Introduction 1
1.1 Overview of Oklahoma Earthquakes and Faulting 1
1.2 Stress Field and Stress State 2
1.3 Triggering Mechanisms 3
1.4 Machine Learning Applications 4
1.5 Structure of the Dissertation 5

References 7

2 Deciphering the stress state of seismogenic faults in Oklahoma and southern Kansas based on an improved stress map 15
2.1 Introduction 15
2.2 Data 19
2.3 Methods 20
2.3.1 Clustering and Fault Mapping 20
2.3.2 Stress Inversion Method 21
2.3.3 Focal Mechanism Tomography (FMT) 22
2.4 Results 25
2.4.1 Clustering and Fault Mapping 25
2.4.2 Stress Field 26
2.4.3 Fault Stress State 27
2.5 Discussion 28
2.5.1 Influence of Pore Pressure on Stress Field and Stress State 28
2.5.2 Stress Tensor Heterogeneity 31
2.5.3 Effect of the Friction Coefficient 32
2.5.4 Seimogenic Faults, Sedimentary Faults, and the Common Tectonic Control 33
2.5.5 Faults with M$\geq$ 5.0 Earthquakes 35
2.5.6 Earthquake Hazard Potential 37

2.6 Conclusions . . . 39
2.7 Supplementary Materials . . . 41

References . . . 44

3 Diverse Stress State Evolution of Induced Earthquake Sequences in Oklahoma Revealed through High-resolution Focal Mechanisms . . . 79
3.1 Introduction . . . 79
3.2 Data . . . 82
3.3 Method . . . 83
3.3.1 Earthquake Relocation and Fault Structure . . . 83
3.3.2 P-wave Polarity . . . 84
3.3.3 Focal Mechanism Inversion . . . 87
3.3.4 Fault Stress State . . . 88
3.4 Results . . . 89
3.4.1 Fault Initiation and Fault Structures . . . 89
3.4.2 Focal Mechanism Results . . . 91
3.4.3 Fault Stress State Results . . . 93
3.5 Discussion . . . 95
3.5.1 Validity of the Polarity-picking Methods . . . 96
3.5.2 Focal Mechanism Uncertainties . . . 97
3.5.3 Potential Earthquake Hazard . . . 98
3.6 Conclusions . . . 100

References . . . 101

4 Coulomb Stress Transfer Influences Fault Reactivation in Areas of Wastewater Injection . . . 122
4.1 Introduction . . . 123
4.2 Data . . . 125
4.2.1 Earthquake Relocations . . . 125
4.2.2 Focal Mechanism Solutions . . . 126
4.3 Fault Interpretation . . . 127
4.4 Fault Reactivation Process . . . 128
4.5 Discussion: Why is the Earthquake Interaction Important? . . . 131
4.5.1 Characteristics in Seismicity Distribution vs the Poisson Process . . . 131
4.5.2 Pore Pressure, Poroelastic Stress, and Coulomb Stress Transfer . . . 133
4.5.3 Role of Fault Zone Structure in Fault Activation Process 134
4.6 Conclusions . . . 135

References . . . 137

5 Forecasting Induced Seismicity in Oklahoma using Machine Learning Methods . . . 145
5.1 Introduction . . . 146
5.2 Data . . . 148
5.2.1 Earthquakes . . . 148
5.2.2 Wastewater Injection . . . 148
5.2.3 Pore Pressure and Poroelastic Stress Data . . . 149
5.3 Methods . . . 149
5.4 Results . . . 152
5.5 Discussion . . . 155
5.5.1 Parameters to Build Input: Grid Size, Prediction Window, Grid Plus . . . 155
5.5.2 Local Stress State and Spatial Heterogeneity . . . 156
5.5.3 Potential Magnitude Forecast . . . 157
5.6 Conclusions . . . 158

References . . . 159

6 Conclusions and Future Work . . . 176
6.1 Summary of Conclusions . . . 176
6.2 Future Directions . . . 178

References . . . 180

LIST OF FIGURES

Figure 1.1: Monthly wastewater injection volumes into Arbuckle Group (blue line) and monthly seismicity (M3.0+) from OGS catalog (red line) in Oklahoma. 13

Figure 1.2: Definition of understress. The Mohr circles represent the stress field, and the black lines represent the fault strength under different pore pressures. The point (σ_0, τ_0) represents the stress state of a fault plane. By definition, understress is calculated as $(\tau_p - \tau_0)/\tau_p$. The background color shows the stress state of faults of various orientations. The parameter understress reflects the criticality of the faults. 14

Figure 2.1: Mapped faults in this study based on earthquake relocations from C. Chen (2016) and Schoenball and Ellsworth (2017b). (a) Earthquake clusters (colored dots) with 10 and more events. Short, black lines show the faults with planarity larger than 0.8 mapped from 30 and more events. Yellow stars show the location of four M≥5 earthquakes in Oklahoma. Thin, black lines are county boundaries in Oklahoma. The long, thick black line is the Nemaha fault from OGS fault database (Marsh & Holland, 2016). (b) The inset map shows the location of the study area. (c) Histogram of strike of faults in (a). (d) Histogram of dip of faults in (a). . . . 55

Figure 2.2: Stress inversion results for Oklahoma and southern Kansas. (a) The bars show the orientation of the maximum horizontal stress (σ_{Hmax}) axis. Green: strike-slip faulting; black: oblique normal faulting. Faulting regime is assigned according to Zoback (1992). The number in the grid shows the corresponding R value. Gray circles are focal mechanisms used in the inversion. Thin, gray lines are county boundaries in Oklahoma. The long, thick black line is the Nemaha fault. (b) Rose diagram of σ_{Hmax} orientation. (c) Cross plot of R value and calculated pore pressure from Langenbruch et al. (2018). The background shows the calculated pore pressure by November 2016 from hydrogeologic models (Langenbruch et al., 2018). 56

Figure 2.3: The stress state of 69 seismogenic faults in a 3D Mohr diagram. The three semicircles represent the stress tensor, and the two straight lines represent the fault strength under hydrostatic fluid pressure and lithostatic pressure under friction coefficient of 0.68. Each circle, colored by understress value, represents the shear and normal stress on a single fault. 57
Figure 2.4: (a) Seismogenic fault map colored by excess pore pressure. Yellow stars show the location of four M≥5 earthquakes in Oklahoma. The dashed lines are county boundaries. The long, thick black line is the Nemaha fault. The background shows the calculated pore pressure by November, 2016 from Langenbruch et al. (2018). (b) Histogram of excess pore pressure on each fault. 58
Figure 2.5: Fault understress values from each stress field scenario. (a) Crossplot of understress values between scenario S1 and S2. (b) Histogram of understress difference between scenario S2 and S1. (c) Crossplot of understress values between scenario S3 and S2. (d) Histogram of understress difference between scenario S3 and S2. 59
Figure 2.6: Dependence of the fault stress state on friction coefficients of 0.4, 0.6, 0.8, and 1.0. The plot schemes are the same as in Figure 2.3. 60
Figure 2.7: Strike comparisons of different categories of faults. (a) Seismogenic faults in the western pressurized section. (b) Mapped sedimentary faults in the western section. (c) Seismogenic faults in the eastern pressurized section. (d) Mapped sedimentary faults in the eastern section. (e) Mapped fractures in Mill Creek and Tishomingo, Oklahoma. The mapped sedimentary faults are from Marsh and Holland (2016). (f) The fault map with the same color scheme as the rose diagrams. The western and eastern pressurized sections are separated by the Nemaha Fault. Yellow dots denote the locations of exposed fractures. Some long N-S trending faults from (Marsh & Holland, 2016) are labeled, including the Nemaha Fault, Wilzetta Fault, Keokuk Fault, and the Galena Township Fault. The figure also labels the geological province in the south of Oklahoma (Northcutt & Campbell, 1996). 61

Figure 2.8: Understress for (a) the focal mechanism solutions of M≥5.0 mainshocks and (b) the hosting seismogenic faults as listed in Table S1. Diamonds: focal mechanism solutions. Circles: seismogenic faults. Both symbols are colored by understress values and scaled with the magnitude of the mainshock. Pr: Prague; Pa: Pawnee; C: Cushing; F: Fairview. 62
Figure 2.9: (a) Scatter plot of the maximum earthquake magnitude and understress value on each fault. The dots are colored by excess pore pressure. (b) Scatter plot of maximum earthquake magnitude and fault length. The dots are colored by understress value. The black line is the empirical relation between magnitude and rupture length from natural earthquakes: M= 4.33+1.49 log(RLD) (M>4.5), RLD–subsurface rupture length (km) (Wells & Coppersmith, 1994). 63
Figure S2.1: Histogram of location difference for common events in the catalog from C. Chen (2016) and Schoenball and Ellsworth (2017b). 64
Figure S2.2: Grids and focal mechanism solutions used in stress inversion. The number of focal mechanisms available is shown in each grid. Gray circles are focal mechanisms used in the inversion. The bars show the orientation of the maximum horizontal stress (σ_{Hmax}) axis from inversion. Green: strike-slip faulting; black: oblique normal faulting. Faulting regime is assigned according to Zoback (1992). Thin, gray lines are county boundaries in Oklahoma. The long, thick black line is the Nemaha fault. 65
Figure S2.3: An example of principal stress and excess pore pressure (required pore pressure increase for fault failure) changes with depth. 66
Figure S2.4: A close-up view (west of north-central Oklahoma) of the clusters. 67
Figure S2.5: A close-up view (east of north-central Oklahoma) of the clusters. 68
Figure S2.6: The uncertainty of (a) σ_{Hmax} orientation and (b) R value estimated from bootstrap resamplings in stress inversion. Small uncertainties of σ_{Hmax} and R value suggest that the heterogeneity of the stress field is well constrained using the high quality focal mechanism solutions. 69

Figure S2.7: (a) Cross plot of understress values and averaged pore pressure on each fault. (b) Cross plot of excess pressure values and averaged pore pressure on each fault. East and west sections are separated by the Nemaha fault. We bin the pore pressure with a bin size of 0.02MPa. The open circles in (a) show the average values of understress which fall in the pore pressure bin of 0.02MPa. The pore pressure was calculated from hydrological model by Langenbruch et al. (2018). 70
Figure S2.8: Required pore pressure calculated by adding uncertainties to fault strike and slip. The red points are the original pore pressure. The blue points are the mean pore pressure by varying fault strike and dip 200 times. The black lines show the range of one standard deviation. 71
Figure S2.9: The magnitude-time plot (blue dots) and cumulative seismicity rate (red lines) of earthquake sequences in (a) Prague, (b) Fairview, (3) Pawnee, and (4) Cushing, Oklahoma. The raw catalog from the Oklahoma Geological Survey is used in the plot. For Fairview and Cushing, the M5 sequences are selected based on temporal clustering of the whole sequence. 72
Figure S2.10: Seismicity distribution in the four M5 clusters. The yellow stars are the mainshock. Events before mainshock are in red, after in gray. The event locations are from Schoenball and Ellsworth (2017b) catalog. 73
Figure S2.11: Definition of understress (Gischig, 2015) and instability (Vavryčuk, 2014). The largest Mohr circle is tangent to the fault strength line under hydrostatic pore pressure based on the assumptions listed in the manuscript. By definition, understress is calculated as $|BA|/|BC|$, and instability is calculated as $|A_1C_1|/|B_1C_1|$. 74
Figure S2.12: Stress state of faults of various orientations represented by (a) understress, and (b) instability. Both parameters can reflect the criticality of the faults. 75
Figure S2.13: Cross plot of understress and instability for 69 seismogenic faults. The bifurcation at high instability values corresponds to the area between the largest and second largest Mohr circles in Figure S2.12. 76

Figure 3.1: Stations (gray triangles) and earthquakes (colored circles in four sequences, 2009–2018) used in this study. The inset figure shows the location of the study region. 106

Figure 3.2: Earthquake relocations for four sequences. The top panels show the map view, and the bottom panels show the cross-section view (along AA′). The events are colored by event origin times. Star: M≥5 event; circles: M≥4 events. The inset diagrams on the map view show the interpretation of the fault structure. The diagrams are not to scale. 107
Figure 3.3: Histogram of strike and dip angles of focal mechanisms for the four sequences. The red lines show the orientation of the main faults interpreted from seismicity. For Cushing and Fairview, the magenta lines show the strike of M≥5 earthquakes. 108
Figure 3.4: Focal mechanism solutions for four sequences. The top panels show the cross-section view along main fault strike, and the bottom panels show the cross-section view (along AA′). The events are colored by fault classes. S: strike-slip, N: normal, R: reverse. Star: M≥5 event; circles: M≥4. . . 109
Figure 3.5: The stress state of individual fault planes in each sequence in 3D Mohr diagram. The three semi-circles represent the stress tensor, and the two straight lines represent the fault strength under hydrostatic fluid pressure and lithostatic pressure assuming a friction coefficient of 0.68. 110
Figure 3.6: Fault stress state for four sequences. The top panels show the cross-section view along strike, and the bottom panels show the cross-section view (along AA′). The events are colored by understress values. Star: M≥5 event; circles: M≥4.111
Figure 3.7: Temporal evolution of the four sequences. The top panel shows the event magnitude with time (black and gray for events with and without inverted focal mechanisms). The middle panel shows the focal mechanism classes with time. The bottom panel shows the stress state of individual fault planes with time. The gray filled events are focal mechanisms of quality D. Star: M≥5 event; circles: M≥4. 112
Figure 3.8: Cross plot of understress and modeled pore pressure for each sequence. The events are colored by event time and scaled by magnitude. The dashed lines denote the understress value of 0.15. The pore pressure data are from Zhai et al. (2019). 113
Figure S3.1: The temporal distribution of focal mechanisms and the number of available stations for Woodward sequence.The reverse events mainly occur at the beginning and end of the sequence. 114

Figure S3.2: Focal mechanism solutions with observed polarities for some selected reverse events in Woodward sequence. 115
Figure S3.3: The faulting type inverted from resampled input dataset for the reverse events in Woodward. 116
Figure S3.4: Distribution of the mean rotation angle between the FMS from complete input dataset and those from 100 resamplings. 117
Figure S3.5: Distribution of the standard deviations of f-value calculated using FMS inverted from 100 resamplings. 118
Figure S3.6: Distribution of the standard deviations of understress value calculated using FMS inverted from 100 resamplings. . . . 119

Figure 4.1: Seismicity, focal mechanisms and fault interpretation in the Woodward cluster. (a) Seismicity distribution in map view. (b) Seismicity distribution in AA′ cross-section view. (c) Focal mechanism distribution in map view using lower hemisphere projection. (d) Focal mechanism distribution in AA′ cross-section view. Red: normal faulting; green: strike-slip faulting; blue: reverse faulting in (c) and (d). (e) Interpreted fault segments (black lines) in map view. (f) Interpreted fault segments (black lines) in AA′ cross-section view. Dashed black lines represent segments where fault geometry is uncertain in (e) and (f). Relocations are from Schoenball and Ellsworth (2017b). 144
Figure 4.2: Time distribution of seismicity in the Woodward cluster. Top: daily seismicity rate (black) and cumulative moment (red) curves. The dashed line denotes the daily seismicity rate of 1.5. Six stages are highlighted based on the date ranges with daily seismicity rate larger than 1.5. Bottom: magnitude distribution along time. The dashed line denotes the magnitude of 3.0. Note that the maximum magnitude in this cluster is smaller than 4.0. Relocations are from C. Chen (2016). 145
Figure 4.3: Spatial distribution of seismicity in the Woodward cluster. For each stage, the top panel shows the map distribution and the bottom panel shows the depth distribution. Red: events in current stage. Gray: events in previous stages. Black: events between last stage and current stage. The receiver faults mapped in Figure 1 are shown under the seismicity. The seismicity migration is denoted by black arrows. Note the bilateral migration of seismicity to the northeast (deeper) and southwest (shallower) in the cluster. Relocations are from C. Chen (2016). 146

Figure 4.4: Coulomb stress transfer calculated from the source faults (Table 4.1b) on the receiver faults (Table 4.1a), μ=0.68. Each panel shows the cumulative Coulomb stress transfer from stage 1 to current stage. The open circles are earthquake locations from next stage. The gray squares represent the source fault patches in each stage. 147
Figure S4.1: Example of a flower structure mapped from a seismic reflection profile. (a) Seismic amplitude cross-section from a 3D seismic reflection data in Osage county, Oklahoma. (b) Interpreted flower structure (red dotted line) from seismic reflection data profile in Figure S4.1a. The seismic data is a post-stack time-migrated 3D volume provided by SpyGlass Energy. 149
Figure S4.2: Focal mechanisms classification diagrams for (a) FMS shallower than 7.7km and (b) FMS deeper than 7.7km. The classification diagram is based on SMT axes plunges by Kagan (2005). SS: strike-slip faulting, R: reverse faulting, and N: normal faulting. The depth of 7.7km separates the deeper strike-slip faults from shallower normal faults. . . . 150
Figure S4.3: Coulomb stress transfer calculated from the source faults 1 and 2 with different locations. The first row shows the magnitude-weighted average locations used in this paper. The second and third row show two randomly selected locations in the corresponding earthquake cloud. 151
Figure S4.4: Histogram of Coulomb stress transfer of all grids on receiver faults (Table 1a) under different friction coefficients. The vertical, solid lines denote the Coulomb stress transfer of 0, 0.01MPa and 0.1MPa. The vertical, dashed lines denote the median value of ΔCFS of all grids on receiver faults. . . . 152
Figure S4.5: 2D density plot of nearest distance T and R for (a) events in the Woodward cluster and (b) events in the cluster with event times replaced by random times from a Poisson process. 153
Figure S4.6: Histogram of Coulomb stress transfer of all grids on receiver faults (Table 1a with strikes of 88°, 80°, and 87° from earthquake cloud) under different friction coefficients. The vertical, solid lines denote the Coulomb stress transfer of 0, 0.01MPa and 0.1MPa. The vertical, dashed lines denote the median value of ΔCFS of all grids on receiver faults. . . . 154

Figure 5.1: Map view of the seismicity and well injections. Triangles: injection wells scaled by the cumulative injection volume from 2010 to 2018. Circles: M>=2.7 earthquakes from 2010 to February 2020, colored by their origin times. The background shows the modeled pore pressure in June, 2015 from Langenbruch et al. (2018). The inset figure shows the location of our study area. 163
Figure 5.2: The forecast earthquake number from the model. The annual forecast is the sum of the monthly prediction in each year. The first panel shows the location of subregions used in next figure. 164
Figure 5.3: Seismicity rate forecast in (a) the whole study area, (b) central Oklahoma, (c) western Oklahoma, (d) Pawnee sequence, (e) Prague sequence, and (f) Fairview sequence. The gray lines are observations, and the blue and red lines are forecasting results from training and test dataset, respectively. The red dashed line shows forecast for year 2019 and 2020, where the injection data are not available yet. The location of the subregions is shown in Figure 5.2. 165
Figure 5.4: Histogram of feature importance from the model. The most important features are pore pressure, poroelastic stress, injection rate, injection volume, and injection depth. For each category of feature, the feature importance of all related features is summed together. 166
Figure S5.1: Seismicity rate forecast using linear regression in (a) the whole study area, (b) central Oklahoma, (c) western Oklahoma, (d) Pawnee sequence, (e) Prague sequence, and (f) Fairview sequence. The gray lines are observations, and the blue and red lines are forecasting results from training and test dataset, respectively. The red dashed line shows forecast for year 2019 and 2020, where the injection data are not available yet. 167
Figure S5.2: Histogram of feature importance for linear model. The feature importance are the sum of the absolute value of the coefficients of the linear model. 168

Figure S5.3: Seismicity rate forecast using only pore pressure and poroelastic stress related features in random forest in (a) the whole study area, (b) central Oklahoma, (c) western Oklahoma, (d) Pawnee sequence, (e) Prague sequence, and (f) Fairview sequence. The gray lines are observations, and the blue and red lines are forecasting results from training and test dataset, respectively. The red dashed line shows forecast for year 2019 and 2020, where the injection data are not available yet. The R^2 score is 0.38. 169
Figure S5.4: Histogram of feature importance for random forest model using only pore pressure and poroelastic stress related features. For each type of feature, the feature importance of all related features is summed together. 170
Figure S5.5: Seismicity rate forecast using only operational parameters in random forest in (a) the whole study area, (b) central Oklahoma, (c) western Oklahoma, (d) Pawnee sequence, (e) Prague sequence, and (f) Fairview sequence. The gray lines are observations, and the blue and red lines are forecasting results from training and test dataset, respectively. The red dashed line shows forecast for year 2019 and 2020, where the injection data are not available yet. The R^2 score is 0.35. 171
Figure S5.6: Histogram of feature importance for random forest model using only operational parameters. For each type of feature, the feature importance of all related features is summed together. 172
Figure S5.7: R^2 score and MSE (mean squared error) variations with grid size. 173
Figure S5.8: R^2 score and MSE variations with prediction window size. 174
Figure S5.9: R^2 score and MSE variations with grid plus parameter. The grid plus parameter is an enlarged area around each grid to search for wells, pore pressure, and poroelastic stress points. 175
Figure S5.10: The importance of region parameter for different grid sizes. 176

LIST OF TABLES

Table S2.1: Focal Mechanisms of $M \geq 5.0$ Earthquakes and Mapped Fault Geometry. 77
Table S2.2: Statistics of large earthquakes on optimally and non-optimally oriented faults. 78

Table S3.1: Quality characterization of inverted focal mechanisms. . . . 120
Table S3.2: Quality characterization results of inverted focal mechanisms. 121
Table S3.3: Polarity result comparison from SVD and ML methods. . . 122

Table 4.1: Fault Parameters Mapped from Seismicity and Focal Mechanism Solutions. 148
Table S4.1: Information Used in the Calculation of Source Fault. 155

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deep and sincere gratitude to my research advisor Prof. Xiaowei Chen for her guidance, motivation, and continuous support of my Ph.D study. Her knowledge, vision, and enthusiasm have helped me carry out my research and inspired me to explore more interesting topics in geophysics. I could not have imagined having a better advisor and mentor during the last five years.

Besides my advisor, I would like to thank the rest of my committee: Prof. Brett Carpenter, Prof. Michael Behm, Dr. Jake Walter, and Prof. Xingru Wu for their encouragement, insightful comments, and hard questions.

My sincere thanks also goes to Dr. Ting Chen and Dr. Lianjie Huang at Los Alamos National Laboratory, for offering me internship and postdoc opportunities in their groups and leading me working on diverse exciting projects.

I thank my dearest friends at my department: Colin Pennington, Pranshu Ratre, Raymond Ng, Deepankar Dangwal, Angie D. Ortega Romo, and Jiewen Zhang for bringing me all the joyful moments. Their help and company have made my Ph.D journey much more smooth and delightful. I also thank all the teachers and staff at my department, especially Rebecca Fay, for their knowledge and help.

Last but not the least, I would like to thank my family for their love and full support throughout my life. Thanks for always believing in me and giving me strength to reach my dreams!

VITA

2020 Ph.D. in Geophysics, University of Oklahoma, Norman

2017 M.S. in Geophysics, University of Oklahoma, Norman

2014 B.S. in Geophysics, University of Science and Technology of China

PUBLICATIONS

Qin, Y., X. Chen, and T. Chen, Diverse stress state evolutions of induced earthquake sequences in Oklahoma revealed through high resolution focal mechanism solutions, in prep.

Qin, Y., T. Chen, and X. Ma, Forecasting induced seismicity in Oklahoma using machine learning methods, in prep.

Qin, Y., Chen, X., Walter, J. I., Haffener, J., Trugman, D. T., Carpenter, B. M., and Kolawole, F. (2019). Deciphering the stress state of seismogenic faults in Oklahoma and Southern Kansas based on an improved stress map. Journal of Geophysical Research: Solid Earth, 124(12), 12920-12934. https://doi.org/10.1029/2019JB018377

Qin, Y., Chen, X., Carpenter, B. M., and Kolawole, F. (2018). Coulomb stress transfer influences fault reactivation in areas of wastewater injection. Geophysical Research Letters, 45(20), 11-059. https://doi.org/10.1029/2018GL079713

ABSTRACT OF THE DISSERTATION

Roles of Fluid Injection, Stress State, Geological Structure, and Earthquake Interaction on Oklahoma Earthquakes

by

Yan Qin

Doctor of Philosophy in Geophysics

University of Oklahoma, Norman, 2020

Xiaowei Chen, Chair

Fault location and geometry are critical considerations in the evaluation of earthquake hazards. In the first part of the dissertation, I map seismogenic faults in Oklahoma and southern Kansas and analyze their stress state. Most of the mapped faults have near vertical planes (planarity$>$ 0.8 and dip$>$ 70°),

and there is a strong correlation between fault length and maximum magnitude on each fault. The fault trends show prominent conjugate sets that strike [55° ∼ 75°] and [105° ∼ 125°]. A comparison with mapped sedimentary faults and basement fractures reveals common tectonic control. I then invert for an improved stress map using high-quality focal mechanisms. The regional stress map shows a gradual transition from oblique-normal faulting in western Oklahoma to dominant strike-slip faulting in central and northern Oklahoma. Stress amplitude ratio shows a strong correlation with pore pressure from hydrogeologic models, suggesting a measurable pore pressure influence on stress patterns. Finally, I assess fault stress state via 3D Mohr circles; a parameter *understress* is used to quantify the level of fault criticality (with 0 meaning critically stressed faults and 1 meaning faults have no resolved shear stress). The results show that 78% of the faults are critically stressed (understress$\leq$ 0.2), while several seismogenic faults are misoriented with high understress ($>$0.4). Fault geometry and local stress fields can be used to evaluate potential seismic hazard, as the largest earthquakes tend to occur on long, critically stressed faults.

In the next study, I focus on four earthquake sequences with different behaviors in Oklahoma: two sequences with complex fault mesh networks and maximum magnitude around 4 (Woodward and Guthrie); two sequences with relatively larger faults and M$\geq$5 earthquakes (Cushing and Fairview). For each of the four sequences, I map the fault structures, invert focal mechanisms, and analyze the stress state of individual events. The results show that the main fault structures in Cushing and Fairview are near-vertical strike-slip faults, and over 70% of the small earthquakes are optimally oriented to the regional stress field. In contrast, Guthrie and Woodward sequences show more complex

structures. In Guthrie, the main fault is vertical at shallow depth and dipping at 70° to the northeast at deeper depth. In Woodward, the main fault is characterized by a flower structure with decreasing dipping angles at shallower depth. The inverted focal mechanisms show a mix of strike-slip faulting and normal faulting in both sequences. The percentage of optimally oriented events is 59% and 47% for Guthrie and Woodward. The results could explain the differences in temporal seismicity evolution in the four sequences. In Cushing and Fairview, more optimally oriented events and higher pore pressure result in maximum magnitudes larger than 5.0. However, in Guthrie and Woodward, fewer optimally oriented events form typical swarm-type sequences without significant change in earthquake magnitude and seismicity rate. My results reflect the heterogeneous fault structures and focal mechanisms in Oklahoma, which possibly influence the seismicity evolution and earthquake hazards in different sequences.

During the above study, the results show that some less- or non-optimally oriented events occur following the optimally oriented events in some sequences, suggesting the influence of processes other than injection, such as earthquake interactions. So, I explore the role of Coulomb stress transfer in fault reactivation in Woodward, Oklahoma. I address this issue by first defining fault segments from earthquake spatiotemporal clustering, then parameterizing the geometries of each segment by combining seismicity and focal mechanisms, and finally calculating Coulomb stress transfer along each fault segment. The results reveal a fault system characterized by a "flower structure" with a strike-slip fault at deeper depth and distributed normal faults at shallower depth. Further, Coulomb stress analysis reveals that the fault reactivation initiates at the fault bend and sequentially migrates to northeast and southwest due

to inter-event stress interaction. The amplitude of Coulomb stress transfer is at least comparable to pore pressure and poroelastic stress changes estimated from fluid injection. Overall, my observations suggest that fault structure and Coulomb stress transfer constitute important factors in seismogenic fault reactivation within areas of wastewater injection.

Other than earthquake interaction, pore pressure increase and poroelastic stress change from injection, are still regarded as the driving mechanisms of seismicity in Oklahoma. In the last study, I use machine learning methods to forecast the seismicity rate in Oklahoma directly from injection parameters and its associated pore pressure and poroelastic stress changes. I divide the study area into uniform grids and search for injection wells, modeled pore pressure and poroelastic stress points, and earthquakes in each grid. The parameters related to injections are features used to forecast earthquake numbers in a Random Forest model. I split the data into training (2010–2016) and test (2017–2020) dataset. The model can forecast the rapid seismicity decrease in recent years. The model also shows that pore pressure and poroelastic stress are the most important features in the forecasting. The findings are consistent with the known mechanisms of induced seismicity. My study suggests that machine learning techniques can be applied to forecasting earthquakes and understanding the physics behind induced seismicity.

Chapter 1

Introduction

1.1 Overview of Oklahoma Earthquakes and Faulting

Oklahoma has experienced drastic seismicity rate changes during the last decade (Figure 1.1). Before 2009, the background seismicity rate was about two M3.0 earthquakes per year (Walter, Ogwari, Thiel, Ferrer, Woelfel, Chang, et al., 2020). The seismicity rate started to increase after 2010, reached the peak during 2015–2016 (901 and 619 M3.0+ earthquakes in 2015 and 2016, respectively), and rapidly decreased due to the reduced wastewater injection. As shown in Figure 1.1, from 2010 to late 2014, state wide disposal rates increased from 39 million bbls per month to 89 million bbls per month. In 2018, the injection rates decreased to the level of 2010. Meanwhile, the observation network has greatly improved, including permanent broadband stations, temporary stations, and nodal arrays from the Oklahoma Geological Survey (OGS), the Incorporated Research Institution for Seismology (IRIS), the

United States Geological Survey (USGS), and private research undertakings (Walter, Ogwari, Thiel, Ferrer, Woelfel, Chang, et al., 2020). The current network provides real-time catalog data and waveform data for various analyses and seismic hazard evaluation.

With the increase of seismicity in Oklahoma, considerate efforts have been devoted to mapping faults. The early mapped faults are mainly large faults separating the geological provinces (Northcutt & Campbell, 1996). Marsh and Holland (2016) compiled a more comprehensive fault map from previous literature and data from the oil and gas industry. Most of the faults are located in the sedimentary layers. Recently in seismology, researchers mapped the reactivated faults based on the alignment of earthquake locations (e.g., Keranen et al., 2013; Yeck et al., 2016; Chen et al., 2017; Schoenball et al., 2018; Qin et al., 2019). A better picture of the active fault system helps with the understanding of potential earthquake hazard and provides information for wastewater injection mitigation strategies.

1.2 Stress Field and Stress State

In the ambient stress field, the optimally oriented faults are more susceptible to fail. Earlier studies from Holland (2013b) and Darold and Holland (2015) identified optimally oriented faults based on probability density functions of fault strikes assuming a uniform maximum horizontal compressional stress orientation of N85°E. Alt and Zoback (2017) analyzed regional stress fields for Texas, Oklahoma, and Kansas combining stress inversions from focal mechanism solutions and wellbore measurements. Walsh and Zoback (2016) simulated the conditional probability of fault slip related to injection-induced

earthquakes by incorporating the uncertainty of the stress tensor, pore pressure, friction coefficient, and fault orientation. In this dissertation, I invert a new stress map using high-quality focal mechanisms from the Oklahoma Geological Survey (OGS).

The stress state of faults is analyzed in 3D Mohr circles using focal mechanism tomography (FMT) (Terakawa et al., 2010). The assumptions include: (1) fault strength is controlled by the Coulomb failure criterion with a constant friction coefficient (Byerlee, 1978); (2) seismic slip occurs in the direction of the resolved shear traction acting on pre-existing faults (Wallace, 1951; Bott, 1959); and (3) seismic slip on optimally oriented faults relative to the regional stress pattern occurs under hydrostatic pressure. To measure the criticality of the faults, I use a parameter *understress* to quantify the difference between the fault stress state and the fault strength (Figure S2.11). Values of understress near 0 imply that the faults are optimally oriented, while values near 1 imply the faults are least optimally oriented.

1.3 Triggering Mechanisms

The basic mechanism driving the seismicity in Oklahoma is pore pressure increase due to wastewater injection. The theory is well established: increased pore pressure in the fault zone will lead to a reduction in the effective normal stress on the fault, thereby reducing fault strength and promoting fault slip (e.g., King Hubbert & Rubey, 1959; Healy et al., 1968; Raleigh et al., 1976). Based on this mechanism, various analyses have been used to study on the relationship between earthquakes and injections. For example, the earthquakes are associated with injections based on spatiotemporal correlations (e.g., Barbour

et al., 2017; Chen et al., 2018; Langenbruch & Zoback, 2016; Weingarten et al., 2015). The fluid diffusion curves show that seismicity migration pattern is controlled by the pressure front(e.g. Haffener et al., 2018; Goebel et al., 2017). Pore pressure change from hydrogeological modeling provides a triggering threshold for the earthquakes (Keranen et al., 2014; Goebel et al., 2017). Based on the modeled pore pressure and poroelastic stress perturbations, the induced seismicity rate is forecasted, and the results fit the observations (Langenbruch et al., 2018; Zhai et al., 2019).

Growing evidences have shown that poroelastic stress can also drive some of the induced seismicity, especially that distant from the injection source (e.g., Segall & Lu, 2015; Deng et al., 2016; Goebel et al., 2017). Fracturing for the unconventional shale formation also induces small earthquakes, especially in the SCOOP/STACK region in Oklahoma (Holland, 2013a; Skoumal et al., 2018). The seismicity usually nucleates in tight clusters in Oklahoma, so earthquake interactions have also been found in multiple areas (Segall & Lu, 2015; Chen et al., 2017; Pennington & Chen, 2017; Qin et al., 2018). Other factors include aseismic creep (Cappa et al., 2019; Eyre et al., 2019) and dynamic triggering (Peña Castro et al., 2019) have also been found to influence the local seismicity. The influences from multiple processes have made the analyses of seismicity in Oklahoma more complex and interesting.

1.4 Machine Learning Applications

With the exponentially growing data volume available in seismology, the application of machine learning techniques has gained much popularity in recent years (e.g., Perol et al., 2018; Ross et al., 2018). The machine learning

techniques have been applied to: earthquake detection and phase picking, earthquake early warning and real-time processing, ground-motion prediction, seismic tomography, and earthquake geodesy (Kong et al., 2019). Perol et al. (2018) first applied convolutional neural network for earthquake detection and location using a single station in Oklahoma. Recent studies (e.g., Zhang et al., 2020) have developed the deep learning systems with multiple stations to locate the earthquakes. Hincks et al. (2018) used an advanced Bayesian network to evaluate the injection parameters and found that injection depth relative to the crystalline basement most strongly correlated with seismic moment release. The OGS developed a Python package, easyQuake, that leveraged a machine-learning driven phase picker, coupled with an associator, to detect and locate earthquakes (in review, Walter, Ogwari, Thiel, Ferrer, & Woelfel, 2020).

Many more studies have been carried out in other regions (Holtzman et al., 2018; Ross et al., 2018). We can take advantage of the new techniques and apply them to the dataset in Oklahoma. Hopefully the new techniques will promote a better understanding of the induced earthquakes.

1.5 Structure of the Dissertation

The following four chapters are reformatted version of papers that are published or in preparation for submission. The objectives for each chapter are summarized as follows:

- Chapter 2 maps the currently active seismogenic faults and stress field at the state scale. The results are published on *Journal of Geophysical Research: Solid Earth*, https://doi.org/10.1029/2019JB018377.

- Chapter 3 focuses on four individual sequences to evaluate the potential seismic hazard based on fault structures, stress state, and pore pressure perturbations.
- Chapter 4 focuses on the Woodward sequence and finds the driving mechanisms for seismicity evolution as earthquake interactions. The results are published on *Geophysical Research Letters*, https://doi.org/10.1029/2018GL079713.
- Chapter 5 is the application of machine learning methods to forecasting seismicity rate based on injection parameters.
- Chapter 6 summarizes the main findings and proposes future work.

References

Alt, R. C., & Zoback, M. D. (2017). In situ stress and active faulting in Oklahoma. *Bulletin of the Seismological Society of America*, *107*(1), 216–228. doi: 10.1785/0120160156

Barbour, A. J., Norbeck, J. H., & Rubinstein, J. L. (2017). The effects of varying injection rates in osage county, oklahoma, on the 2016 m w 5.8 pawnee earthquake. *Seismological Research Letters*, *88*(4), 1040–1053.

Bott, M. H. P. (1959, mar). The mechanics of oblique slip faulting. *Geological Magazine*, *96*(2), 109–117. doi: 10.1017/S0016756800059987

Byerlee, J. (1978). Friction of rocks. *Pure and Applied Geophysics PAGEOPH*, *116*(4-5), 615–626. doi: 10.1007/BF00876528

Cappa, F., Scuderi, M. M., Collettini, C., Guglielmi, Y., & Avouac, J.-P. (2019). Stabilization of fault slip by fluid injection in the laboratory and in situ. *Science advances*, *5*(3), eaau4065. doi: 10.1126/sciadv.aau4065

Chen, X., Haffener, J., Goebel, T. H., Meng, X., Peng, Z., & Chang, J. C. (2018). Temporal correlation between seismic moment and injection volume for an induced earthquake sequence in central oklahoma. *Journal of Geophysical Research: Solid Earth*, *123*(4), 3047–3064.

Chen, X., Nakata, N., Pennington, C., Haffener, J., Chang, J. C., He, X., ... Walter, J. I. (2017). The pawnee earthquake as a result of the interplay

among injection, faults and foreshocks. *Scientific reports*, *7*(1), 1–18.

Darold, A., & Holland, A. (2015). Preliminary oklahoma optimal fault orientations. *Oklahoma Geol. Surv. Open File Rep., OF4*.

Deng, K., Liu, Y., & Harrington, R. M. (2016). Poroelastic stress triggering of the December 2013 Crooked Lake, Alberta, induced seismicity sequence. *Geophysical Research Letters*, *43*(16), 8482–8491. doi: 10.1002/2016GL070421

Eyre, T. S., Eaton, D. W., Garagash, D. I., Zecevic, M., Venieri, M., Weir, R., & Lawton, D. C. (2019). The role of aseismic slip in hydraulic fracturing–induced seismicity. *Science advances*, *5*(8), eaav7172. doi: 10.1126/sciadv.aav7172

Goebel, T. H., Weingarten, M., Chen, X., Haffener, J., & Brodsky, E. E. (2017). The 2016 Mw5.1 Fairview, Oklahoma earthquakes: Evidence for long-range poroelastic triggering at >40 km from fluid disposal wells. *Earth and Planetary Science Letters*, *472*, 50–61. doi: 10.1016/j.epsl.2017.05.011

Haffener, J., Chen, X., & Murray, K. (2018). Multiscale analysis of spatiotemporal relationship between injection and seismicity in oklahoma. *Journal of Geophysical Research: Solid Earth*, *123*(10), 8711–8731.

Healy, J., Rubey, W., Griggs, D., & Raleigh, C. (1968). The denver earthquakes. *Science*, *161*(3848), 1301–1310. doi: 10.1126/science.161.3848.1301

Hincks, T., Aspinall, W., Cooke, R., & Gernon, T. (2018). Oklahoma's induced seismicity strongly linked to wastewater injection depth. *Science*, *359*(6381), 1251–1255.

Holland, A. A. (2013a). Earthquakes triggered by hydraulic fracturing in south-central oklahoma. *Bulletin of the Seismological Society of America*, *103*(3), 1784–1792.

Holland, A. A. (2013b). Optimal fault orientations within oklahoma. *Seismological Research Letters*, *84*(5), 876–890. doi: 10.1785/0220120153

Holtzman, B. K., Paté, A., Paisley, J., Waldhauser, F., & Repetto, D. (2018). Machine learning reveals cyclic changes in seismic source spectra in geysers geothermal field. *Science advances*, *4*(5), eaao2929.

Keranen, K. M., Savage, H. M., Abers, G. A., & Cochran, E. S. (2013). Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 Mw 5.7 earthquake sequence. *Geology*, *41*(6), 699–702. doi: 10.1130/G34045.1

Keranen, K. M., Weingarten, M., Abers, G. A., Bekins, B. A., & 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

King Hubbert, M., & Rubey, W. W. (1959). Role of fluid pressure in mechanics of overthrust faulting: I. Mechanics of fluid-filled porous solids and its application to overthrust faulting. *Bulletin of the Geological Society of America*, *70*(2), 115–166. doi: 10.1130/0016-7606(1959)70[115:ROFPIM]2.0.CO;2

Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., & Gerstoft, P. (2019). Machine learning in seismology: Turning data into insights. *Seismological Research Letters*, *90*(1), 3–14.

Langenbruch, C., Weingarten, M., & Zoback, M. D. (2018). Physics-based forecasting of man-made earthquake hazards in oklahoma and kansas. *Nature communications*, *9*(1), 1–10. doi: 10.1038/s41467-018-06167-4

Langenbruch, C., & Zoback, M. D. (2016). How will induced seismicity in oklahoma respond to decreased saltwater injection rates? *Science*

advances, *2*(11), e1601542.

Marsh, S., & Holland, A. (2016). Comprehensive fault database and interpretive fault map of oklahoma. *Oklahoma Geol. Surv. Open-File Rep. OF2-2016, Oklahoma Geological Survey, Norman, OK*.

Northcutt, R. A., & Campbell, J. A. (1996). Geologic provinces of oklahoma.

Peña Castro, A., Dougherty, S. L., Harrington, R., & Cochran, E. S. (2019). Delayed dynamic triggering of disposal-induced earthquakes observed by a dense array in northern oklahoma. *Journal of Geophysical Research: Solid Earth*, *124*(4), 3766–3781.

Pennington, C., & Chen, X. (2017). Coulomb stress interactions during the m w 5.8 pawnee sequence. *Seismological Research Letters*, *88*(4), 1024–1031.

Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, *4*(2), e1700578.

Qin, Y., Chen, X., Carpenter, B. M., & Kolawole, F. (2018). Coulomb stress transfer influences fault reactivation in areas of wastewater injection. *Geophysical Research Letters*, *45*(20), 11–059.

Qin, Y., Chen, X., Walter, J. I., Haffener, J., Trugman, D. T., Carpenter, B. M., ... Kolawole, F. (2019). Deciphering the stress state of seismogenic faults in oklahoma and southern kansas based on an improved stress map. *Journal of Geophysical Research: Solid Earth*, *124*(12), 12920–12934.

Raleigh, C. B., Healy, J. H., & Bredehoeft, J. D. (1976). An experiment in earthquake control at Rangely, Colorado. *Science*, *191*(4233), 1230–1237. doi: 10.1126/science.191.4233.1230

Ross, Z. E., Meier, M.-A., & Hauksson, E. (2018). P wave arrival picking and first-motion polarity determination with deep learning. *Journal of Geophysical Research: Solid Earth*, *123*(6), 5120–5129.

Schoenball, M., Walsh, F. R., Weingarten, M., & Ellsworth, W. L. (2018). How faults wake up: the guthrie-langston, oklahoma earthquakes. *The Leading Edge*, *37*(2), 100–106.

Segall, P., & Lu, S. (2015). Injection-induced seismicity: Poroelastic and earthquake nucleation effects. *Journal of Geophysical Research: Solid Earth*, *120*(7), 5082–5103.

Skoumal, R. J., Ries, R., Brudzinski, M. R., Barbour, A. J., & Currie, B. S. (2018). Earthquakes induced by hydraulic fracturing are pervasive in oklahoma. *Journal of Geophysical Research: Solid Earth*, *123*(12), 10–918.

Terakawa, T., Zoporowski, A., Galvan, B., & Miller, S. A. (2010). High-pressure fluid at hypocentral depths in the L'Aquila region inferred from earthquake focal mechanisms. *Geology*, *38*(11), 995–998. doi: 10.1130/G31457.1

Wallace, R. E. (1951). Geometry of shearing stress and relation to faulting. *The Journal of Geology*, *59*(2), 118–130. doi: 10.1086/625831

Walsh, F. R., & Zoback, M. D. (2016). Probabilistic assessment of potential fault slip related to injectioninduced earthquakes: Application to north-central Oklahoma, USA. *Geology*, *44*(12), 991–994. doi: 10.1130/G38275.1

Walter, J. I., Ogwari, P., Thiel, A., Ferrer, F., & Woelfel, I. (2020). easyquake: Putting machine learning to work for your regional seismic network or local earthquake study. *Seismological Research Letters*.

Walter, J. I., Ogwari, P., Thiel, A., Ferrer, F., Woelfel, I., Chang, J. C., ... Holland, A. A. (2020). The oklahoma geological survey statewide seismic network. *Seismological Research Letters*, *91*(2A), 611–621.

Weingarten, M., Ge, S., Godt, J. W., Bekins, B. A., & Rubinstein, J. L. (2015).

High-rate injection is associated with the increase in us mid-continent seismicity. *Science*, *348*(6241), 1336–1340.

Yeck, W. L., Weingarten, M., Benz, H. M., McNamara, D. E., Bergman, E., Herrmann, R., ... Earle, P. S. (2016). Far-field pressurization likely caused one of the largest injection induced earthquakes by reactivating a large preexisting basement fault structure. *Geophysical Research Letters*, *43*(19), 10–198.

Zhai, G., Shirzaei, M., Manga, M., & Chen, X. (2019). Pore-pressure diffusion, enhanced by poroelastic stresses, controls induced seismicity in oklahoma. *Proceedings of the National Academy of Sciences*, *116*(33), 16228–16233.

Zhang, X., Zhang, J., Yuan, C., Liu, S., Chen, Z., & Li, W. (2020). Locating induced earthquakes with a network of seismic stations in oklahoma via a deep learning method. *Scientific reports*, *10*(1), 1–12.

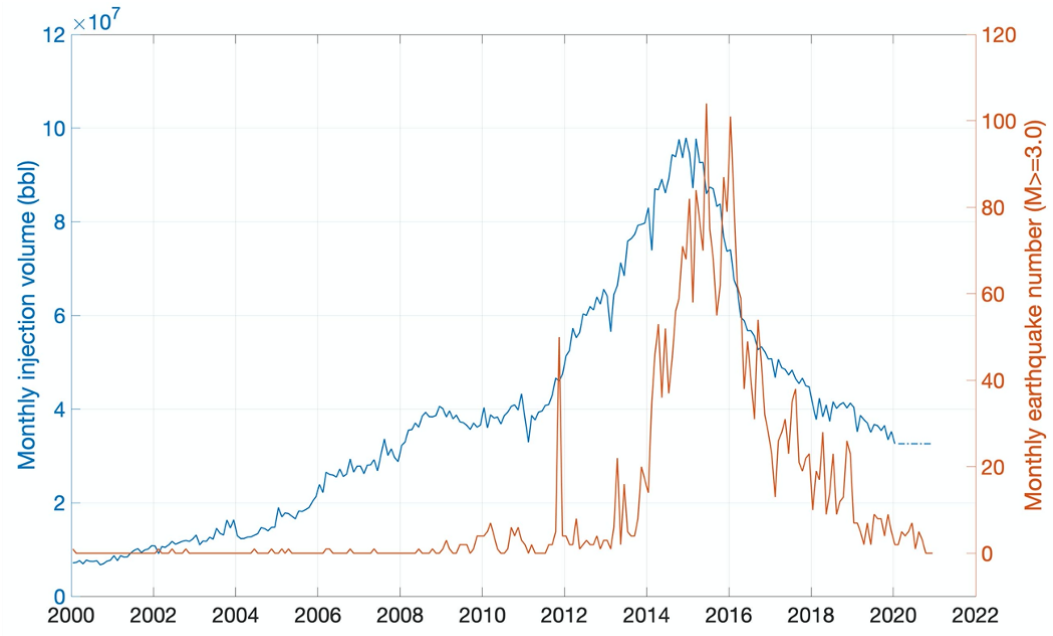

Figure 1.1: Monthly wastewater injection volumes into Arbuckle Group (blue line) and monthly seismicity (M3.0+) from OGS catalog (red line) in Oklahoma.

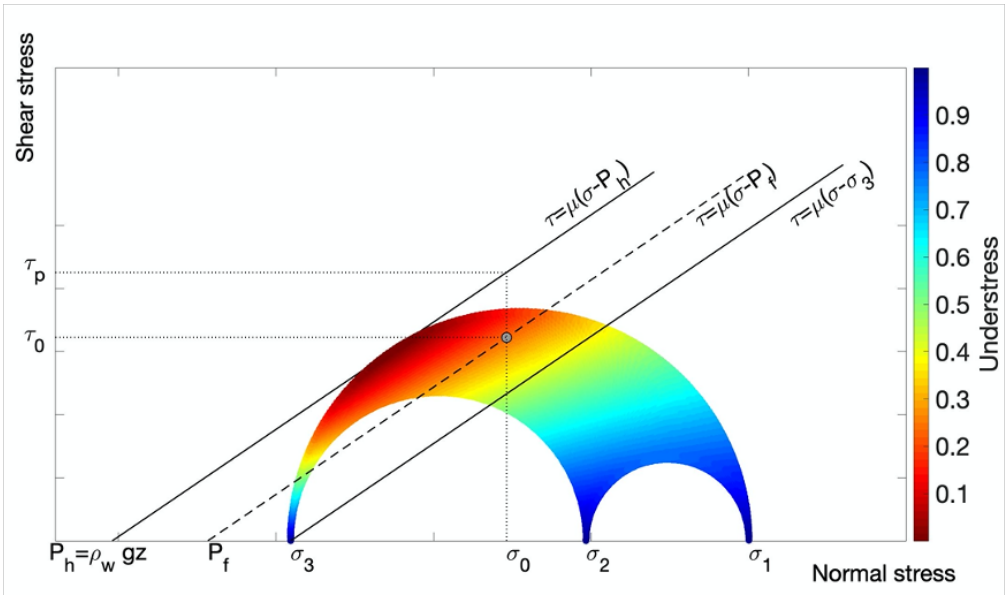

Figure 1.2: Definition of understress. The Mohr circles represent the stress field, and the black lines represent the fault strength under different pore pressures. The point (σ_0, τ_0) represents the stress state of a fault plane. By definition, understress is calculated as $(\tau_p - \tau_0)/\tau_p$. The background color shows the stress state of faults of various orientations. The parameter understress reflects the criticality of the faults.

Chapter 2

Deciphering the stress state of seismogenic faults in Oklahoma and southern Kansas based on an improved stress map

2.1 Introduction

The recent increase in seismicity in Oklahoma has been associated with wastewater injection (e.g., Ellsworth, 2013; Keranen et al., 2014; W. Yeck et al., 2016). The basic mechanism driving this seismicity is well established: increased pore pressure in the fault zone will lead to a reduction in the effective normal stress on the fault, thereby reducing fault strength and promoting fault slip (e.g., King Hubbert & Rubey, 1959; Healy et al., 1968; Raleigh et al., 1976). Based on this mechanism, the orientation of the faults and the regional stress

field are crucial to assess the likelihood of reactivation. Faults that are optimally oriented with respect to the regional stress field will be readily reactivated, while faults that are nonoptimally oriented will require a much larger pore pressure increase to slip. Holland (2013) and Darold and Holland (2015) tried to differentiate optimally oriented faults from nonoptimally oriented faults based on probability density functions of fault strikes relative to a uniform maximum horizontal compressional stress orientation of N85°E. (Alt & Zoback, 2017) analyzed regional stress fields for Texas, Oklahoma, and Kansas combining stress inversions from focal mechanism solutions and wellbore measurements. Walsh and Zoback (2016) simulated the conditional probability of fault slip related to injection-induced earthquakes by incorporating the uncertainty of the stress tensor, pore pressure, friction coefficient, and fault orientation.

While previous studies have provided overall knowledge of the ambient stress field in Oklahoma and southern Kansas, a high-resolution stress map is needed to systematically assess fault criticality under the local stress field. This is especially important with the growing evidence that poroelastic stress (e.g., Segall & Lu, 2015; Deng et al., 2016; Barbour et al., 2017; Goebel et al., 2017) and aseismic creep propagation (e.g., Cappa et al., 2019; Eyre et al., 2019) could, in part, drive some of the induced seismicity. In this study, we develop a stress map with relatively high spatial resolution using a suite of 2047 focal mechanism solutions obtained from Oklahoma and southern Kansas, allowing for more precise quantitative analysis of the fault stress state.

For a given background stress field, a complete knowledge of the pre-existing fault system is critical to evaluate the induced earthquake hazard (W. Yeck et al., 2016; Levandowski, Weingarten, & Walsh III, 2018). Oklahoma

and southern Kansas are located in the Precambrian (∼1.4 Ga) Southern Granite-Rhyolite Province of the Mid-Continent U.S. craton (Denison et al., 1987). This granitic basement hosts most of the current seismicity (Schoenball & Ellsworth, 2017b; Kolawole, Johnston, et al., 2019). This basement is characterized by a structural fabric with dominant discontinuity zones that trend NW-SE, NE-SW, and a minor N-S set (Kolawole, Johnston, et al., 2019). This structural fabric resulted from the multiphase Proterozoic contractional and extensional deformation events that affected the central U.S., e.g., development of Granite-Rhyolite Provinces, and the Mid-Continent Rift (Bickford et al., 2015; Whitmeyer & Karlstrom, 2007). The fault database of Oklahoma (Marsh & Holland, 2016) shows several fault segments and large (>50 km-long) ∼N-S-trending faults within the north-central region. Recent 3D seismic data (Liao et al., 2017; Chopra et al., 2018; Kolawole, Carpenter, et al., 2019) reveal that these large N-S faults and the associated secondary splays are basement-rooted. The Oklahoma Geological Survey (OGS) fault database (Marsh & Holland, 2016) largely contains sedimentary faults that represent (1) reactivation and propagation of the Precambrian basement structural trends (NW, NE, and N-S) into the sedimentary cover (Kolawole, Carpenter, et al., 2019), and (2) additional deformation by the development of pervasive R-shears that splay outwards from the major right-lateral N-S faults (Liao et al., 2017; Chopra et al., 2018; Kolawole, Carpenter, et al., 2019).

The majority of earthquakes in Oklahoma do not occur on currently mapped faults in the OGS database (e.g., Alt & Zoback, 2017; Keranen et al., 2013; Goebel et al., 2017; W. Yeck et al., 2016). The linear trends of seismicity (Figure 2.1) suggest that most earthquakes occur on basement faults that are likely unmapped in the current fault database (Skoumal et al., 2019; Schoenball

& Ellsworth, 2017a). Several studies have used the spatial distribution of seismicity to map fault segments in Oklahoma. For example, splays of the Wilzetta fault were mapped from 2011 Mw5.7 Prague earthquake sequence (Keranen et al., 2013); the extension of a mapped fault segment was delineated from the 2016 Mw5.1 Fairview earthquake sequence (W. Yeck et al., 2016); and the Sooner Lake Fault as the conjugate fault of the mapped Labette Fault was mapped from 2016 Mw5.8 Pawnee earthquake sequence (X. Chen et al., 2017).

In this study, we systematically map the reactivated faults (herein referred to as seismogenic faults) and analyze the fault criticality with an improved knowledge of the local stress field. First, we characterize the geometry of seismogenic faults in Oklahoma and southern Kansas based on earthquake clustering. Second, we use a high quality catalog of focal mechanism solutions to perform a detailed stress inversion. Then we assess the stress state of individual faults with in-situ 3D Mohr circles and evaluate the influence of fault criticality on fault reactivation. Finally, we compare seismogenic faults with mapped sedimentary faults from different sub-regions in Oklahoma and fracture systems in outcrops of the seismogenic basement. Moreover, the stress field and fault stress state are compared with pore pressure from hydrogeologic models to further understand the influence of wastewater injection on fault reactivation. These results help to provide a comprehensive understanding of the roles of pre-existing faulting, fluid injection, and stress state on fault reactivation and potential earthquake hazard.

2.2 Data

High-precision earthquake relocations can reveal tightly clustered seismicity patterns and help identify in-situ fault locations and orientations. In this study, we use the relocated catalog from C. Chen (2016) for pre-2013 events and Schoenball and Ellsworth (2017b) for later events to map seismogenic faults. The catalog in C. Chen (2016) uses a 3D velocity model and the double-difference method (Waldhauser, 2001) with catalog differential times to relocate earthquakes. The catalog in Schoenball and Ellsworth (2017b) uses a 1D velocity model and the double-difference method with differential times derived from waveform cross-correlation to relocate earthquakes from 2013 to 2017, and has high relative location precision (50 m horizontally and 200 m vertically) with the inclusion of industry networks and higher precision of differential times. The magnitude of completeness for catalogs from C. Chen (2016) and Schoenball and Ellsworth (2017b) is 2.7 and 2.8, respectively. We verify that for 13512 common earthquakes from C. Chen (2016) and Schoenball and Ellsworth (2017b), the median location difference is 0.42km (Figure S2.1), which mainly comes from some systematical shift of absolute locations due to the difference in velocity model but does not affect relative locations within clusters. We use C. Chen (2016) catalog mainly for the Prague fault with M5.7 earthquake in 2011. Other pre-2013 faults from C. Chen (2016) do not meet the minimum number requirement of 30 and are not included.

For focal mechanism solutions in Oklahoma, we select 1823 focal mechanism solutions of A and B quality in the catalog provided by OGS (Walter et al., 2019) from January 2010 to August 2018, which are computed via HASH (Hardebeck & Shearer, 2008) program with at least 8 routinely picked first

motion polarities. The selected focal mechanism solutions have an average RMS fault plane uncertainty less than 35° and a station distribution ratio larger than 0.4. For southern Kansas, we apply the same criteria and compute 224 A and B quality focal mechanism solutions using the first motion polarities in the HASH program. The focal mechanism solutions in southern Kansas are consistent with Rubinstein et al. (2018). In total, we have 2047 focal mechanism solutions for stress inversion.

2.3 Methods

2.3.1 Clustering and Fault Mapping

We use a hierarchical clustering program in MATLAB to cluster the earthquakes based on the epicenter of the relocations. In this method, the events are linked based on the nearest distance between each event pair, and a distance cutoff of 0.46km is selected by trail and error to group events with distance smaller than the cutoff into a cluster. The program identifies 84 clusters with more than 30 events (Figure 2.1). The clustering results show similar fault trends as in Schoenball and Ellsworth (2017a) from a different clustering method. We use a relatively longer distance cutoff in the clustering process, and some clusters include several trends of events close to each other, which are then separated manually to calculate the fault geometries.

For each cluster and some manually separated subclusters with 30 or more events, we use principal component analysis (Vidale & Shearer, 2006) to fit a fault plane. First, a 3×3 covariance matrix D from the earthquake hypocenters is calculated, and the eigenvalues ($\lambda_1 \geq \lambda_2 \geq \lambda_3$) and the corresponding

eigenvectors U_1, U_2 and U_3 of D define the principal axes of rotational inertia for the points in each cluster. The first two eigenvectors U_1, U_2 represent the surface of the fault plane, and U_3 is normal to the fault plane. The fault strike and dip angle are calculated from the direction of the normal vector U_3. The planarity of seismicity hypocenters is defined as $1 - \lambda_3/\lambda_2$ (Vidale & Shearer, 2006). At its extremes, a planarity of 1 indicates a perfectly planar shape and a planarity of 0 indicates a nearly spherical cloud of seismicity (Vidale & Shearer, 2006). 69 out of 95 planarity values (73%) are larger than 0.8, suggesting most clusters occur on well-defined fault plane. Since most of the seismogenic faults can be fit with a fault plane, the fault length is estimated manually based on the seismicity extension along the axis of the largest eigenvector.

2.3.2 Stress Inversion Method

To obtain a detailed in-situ stress field, we use the MSATSI software package (Martínez-Garzón et al., 2014) to invert the stress field from earthquake focal mechanism solutions. The MSATSI software is a MATLAB wrapper of the SATSI (Hardebeck & Michael, 2006) based on the inversion from Michael (1984). The inversion relies on three assumptions: (1) the stress field is homogeneous over the spatial and temporal extent of the events in each grid, (2) the focal mechanism solutions are adequately diverse, such as the RMS angular difference from the average mechanism in each grid of at least $\sim 40^\circ - 45^\circ$(Hardebeck & Hauksson, 2001) to constrain the solution, and (3) seismic slip occurs in the direction of the resolved shear traction acting on pre-existing faults. With 2047 focal mechanism solutions for Oklahoma and southern Kansas, the study area is first gridded with 0.4° by 0.4°, and if 100 or more events are in one

grid, the grid is then subdivided into two or four evenly spaced subgrids in latitude and longitude as long as there are still more than 50 events in each subgrid. This method covers as much area as possible in the inversion and ensures adequate variety to constrain the solution (Martínez-Garzón, Ben-Zion, et al., 2016) in each grid. As a result, the study area is separated into 24 grids, and a damped inversion is performed on those grids. A map with the number of focal mechanisms in each grid is shown in Figure S2.2. The inversion results include the orientations of the three principal stress axes and a measure of their relative amplitudes R,

$$R = \frac{\sigma_1 - \sigma_2}{\sigma_1 - \sigma_3} \tag{2.1}$$

where σ_1, σ_2, σ_3 are the maximum, intermediate, and minimum principal stresses, respectively. The uncertainties of the inversion results are estimated by 1000 bootstrap resamplings of the focal mechanism solutions associated within each grid.

2.3.3 Focal Mechanism Tomography (FMT)

The focal mechanism tomography (FMT) technique was developed to estimate the fluid pore pressure field from earthquake focal mechanism solutions under a given stress field (Terakawa et al., 2010). In this study, we adopt the assumptions in Terakawa et al. (2010) to convert the relative stress amplitude to a 3D stress tensor and use the local stress tensor to evaluate the stress state of individual faults. The assumptions are: (1) fault strength is controlled by the Coulomb failure criterion with a constant friction coefficient (Byerlee, 1978), (2) seismic slip occurs in the direction of the resolved shear traction

acting on pre-existing faults (Wallace, 1951; Bott, 1959), and (3) seismic slip on optimally oriented faults relative to the regional stress pattern occurs under hydrostatic pressure. Based on these assumptions, we have:

$$\sigma_1(\sqrt{\mu^2+1}-\mu)-\sigma_3(\sqrt{\mu^2+1}+\mu)=2C-2\mu P_w \tag{2.2}$$

where σ_1 and σ_3 are the maximum and minimum principal stresses, C is the cohesion of fault and assumed as zero in the following calculation, P_w is the hydrostatic pressure at depth, and μ is the friction coefficient without pore pressure influence, assumed to be 0.68 in the analysis. The friction coefficient of 0.68 is based on the average value of lab results of basement rock samples in Oklahoma by Kolawole, Johnston, et al. (2019). The choice of friction coefficient is also consistent with the estimated value of 0.65 using STRESSINVERSE program by Vavryčuk (2014). A sensitivity test of friction coefficient is performed in the discussion section. The derivation of Equation 2.8 is shown in supplemental material.

We further assume that the vertical stress is the weight of overburden,

$$\sigma_v=\rho g z \tag{2.3}$$

where ρ is the rock density, $\rho = 2540\text{kg/m}^3$ (Terakawa et al., 2010), g is the acceleration due to gravity, and z is the depth. The stress inversion results show both strike-slip faulting regime and oblique normal faulting regime, so instead of assuming the vertical stress is the intermediate principal stress, it is strictly derived from the principal stress tensor, $\sigma_v = \sigma_v(\sigma_1, \sigma_2, \sigma_3)$ using the orientations of the principal stresses. The derivation is shown in the

supplemental material.

The ratio of principal stresses from the stress inversion (Equation 2.1) is also incorporated to solve for the intermediate stress amplitude (Quinones et al., 2018). By solving equations 2.1 – 2.3, we can get the stress amplitude for each grid. With the stress tensor known, we compute shear and normal stress on faults, project them onto Mohr circle, and calculate the required pore pressure for fault failure. To project all faults onto the same 3D Mohr circle, we keep the local stress orientations of each grid and calculate a uniform stress amplitude by averaging over all grids. We will justify the use of uniform stress amplitude by comparing the results from uniform and nonuniform stress field in the discussion. The stress amplitudes and fluid pore pressure calculated from the above assumptions are proportional to the depth (Figure S2.3). As the catalogs we use have relatively large depth uncertainty, the depth of the mapped faults is not well resolved. We introduce a normalized parameter understress to eliminate the depth dependence of the fault stress state following Gischig (2015),

$$understress = (\tau_p - \tau_0)/\tau_p \tag{2.4}$$

where τ_0 is shear stress on the fault calculated from the fault geometry and stress orientations, and τ_p is shear stress at which slip initiates based on the Coulomb failure criterion under hydrostatic pore pressure. Since both τ_0 and τ_p increase linearly with depth, the defined parameter understress is independent of depth. The understress can be used to quantitatively measure fault criticality relative to local stress field. Values of understress near 0 imply that the faults are critically stressed, while values near 1 imply negligible resolved shear stress

applied on the fault, and the fault is least favorably oriented. We also calculate the parameter of excess pore pressure, which is defined as the required pore pressure increase above hydrostatic pressure for fault failure according to Mohr circle.

2.4 Results

2.4.1 Clustering and Fault Mapping

We identify 84 clusters with 30 or more events and pick 95 fault segments (some clusters are manually separated based on visual inspection). A subset of 69 faults with planarity larger than 0.8 are selected and used in the following analysis (see Figure 2.1a, Figure S2.4 and S2.5 for close-up views of the clusters). The newly mapped faults from several large earthquake sequences are consistent with previous studies (e.g., Keranen et al., 2013; W. Yeck et al., 2016; W. L. Yeck et al., 2017). Specifically, we find that the main fault in Prague is a splay of the Wilzetta fault with an azimuth of 55° and a dip angle of 86°; the Fairview fault is as an extension of a mapped fault to the southeast; and a conjugate fault pattern is delineated in the Pawnee area. The west-northwest and east-northeast trends of faults in southern Kansas are in agreement with the observations in Rubinstein et al. (2018). Most of the faults in Oklahoma are distributed in the central and northern pressurized regions (Walsh & Zoback, 2016; Skoumal et al., 2019). The distributions of strike and dip angles for these faults are shown in Figure 2.1c and 2.1d. The strike angle is mainly distributed in the ranges of $[55 \sim 75°]$ and $[105 \sim 125°]$, which form conjugate patterns relative to a maximum horizontal compression stress

orientation of N85°E. The majority (> 80%) of seismogenic faults are steeply dipping with a dip angle larger than 70°. Four faults show dip angle smaller than 35°. Those faults are mapped from fewer than 45 relatively scattered events, so the fault geometries are likely not well constrained.

2.4.2 Stress Field

Stress inversion provides a stress field with higher spatial resolution compared to previous studies in Oklahoma. Figure 2.2 shows the map view of maximum horizontal compressive stress (σ_{Hmax}) orientations colored by faulting type. Central Oklahoma is mostly in a strike-slip faulting regime (green bars), whereas north and northwest Oklahoma show a transition from strike-slip to oblique normal faulting regime (black bars). The dominant orientation of σ_{Hmax} is $80° \sim 90°$ (Figure 2.2b). Those observations are consistent with previous studies (e.g., Marsh & Holland, 2016; Qi, 2016; Walsh & Zoback, 2016; Levandowski, Herrmann, et al., 2018). The stress field in southern Kansas is characterized by strike-slip faulting with σ_{Hmax} of $75 \sim 82°$, which are consistent with Rubinstein et al. (2018).

The stress amplitude ratio R also shows spatial variations.The stress field in northern Oklahoma and southern Kansas show smaller R values than other areas, which might indicate the influence of pore pressure. The study area is separated into two major pressure zones using the Nemaha fault as a pressure boundary (Haffener et al., 2018), referred to as the eastern and western pressure zones. We obtain the pore pressure for each grid by averaging pressure values from Langenbruch et al. (2018) at the median occurrence time from all earthquakes within each grid. Using pore pressure values at all grids,

we calculate an average R value from all R values within each pressure bin of 0.04 MPa. The result is shown in Figure 2.2c. For the area west of Nemaha Fault Zone, we obtain negative correlation between R value and pore pressure. A similar relationship between pore pressure and R value has been observed at Geysers geothermal field (Martínez-Garzón, Vavryčuk, et al., 2016). However, the eastern Oklahoma region does not show a clear relationship between R value and pore pressure.

The uncertainties of the σ_{Hmax} orientation and R value are estimated from bootstrap resamplings (Figure S2.6). The highest uncertainty of σ_{Hmax} (defined as one standard deviation) is less than 2°. The uncertainty in the R value is less than 0.05. The inversion results and their small uncertainties suggest that the heterogeneity of the stress field is well constrained using the high quality focal mechanism solutions.

2.4.3 Fault Stress State

Using the FMT method, we first calculate the uniform principal stress amplitude gradients as $\sigma_1 = 30.0$ MPa/km, $\sigma_2 = 24.8$ MPa/km, $\sigma_3 = 15.5$ MPa/km under $\mu = 0.68$. Based on the fault orientation and regional stress field, the shear and normal stress on seismogenic faults are calculated and plotted on 3D Mohr circle. The understress parameter on each fault is determined by equation 3.4. As shown in Figure 2.3, each fault is projected onto a 3D Mohr circle as a point colored by its understress value. Most faults (78%) are close to failure limit of the hydrostatic fault strength with understress smaller than 0.2. Four faults with small dip angle $< 35°$ show large understress (> 0.5). This is possibly due to large uncertainties of the dip angle, and the indication of tensile

failure (fluid pressure beyond σ_3) is actually an artifact. Other than that, there are still several non-optimally oriented faults being reactivated, which might occur at step-overs or rotations at different segments of the optimally-oriented fault and result from either high pore pressure increase or other factors, e.g., static stress change from earthquakes on the main fault.

The required pore pressure to induce failure on each fault is calculated from FMT and shown in Figure 2.4. The median and mean excess pore pressure (above hydrostatic pore pressure) is 2.7 and 6.9 MPa, respectively under an assumption of fault depth of 5km. The uniform depth is chose because the depth of seismogenic faults is not well constrained, and the earthquakes have an average depth of 5km. The required pore pressure increases are consistent with the estimated pore pressure using similar geomechanical analysis for multiple induced clusters Texas (Quinones et al., 2018; Snee & Zoback, 2016). The observed pore pressure range is also consistent with the findings in the Geyser geothermal field in Martínez-Garzón, Kwiatek, et al. (2016), where faults with a broad range of orientations are activated by fluid injection, and the misoriented faults are mostly activated during high injection rates in proximity to the injection wells by an estimated pore pressure increase of $\sim$ 10MPa.

2.5 Discussion

2.5.1 Influence of Pore Pressure on Stress Field and Stress State

In the above analysis, we only consider the pore pressure increase in fault reactivation. The observed negative relationship between R and pore pressure

in western pressure zone in Figure 2.2c could possibly reflect the poroelastic effects by injection. Altmann et al. (2014) gives the analytical solutions to poroelastic equations in 3D isotropic, homogeneous space. In strike-slip regime along vertical direction, pore pressure increase ΔP induces effective stress amplitude changes of $-\frac{2}{3}\Delta P$, $-\frac{1}{3}\Delta P$, and $-\frac{2}{3}\Delta P$ in σ_1, σ_2, σ_3 orientations, respectively. The pore pressure brings the σ_1 and σ_2 closer and results in a smaller R value. Martínez-Garzón et al. (2013) has observed that stress perturbation due to fluid injection decreases over time with repeated injections. The lack of correlation between R and pore pressure in the eastern section is likely due to the overall higher pore pressure from the longer injection period (e.g., Keranen et al., 2014).

To study the relationship between pore pressure and fault reactivation, we compare our results to the modeled pore pressure from Langenbruch et al. (2018). The modeled pore pressure map is overlain by seismogenic faults colored by excess pore pressure computed in this study in Figure 2.4. As a qualitative first-order observation, the faults that are misoriented and require a relatively larger pore pressure increase are distributed close to the higher pore pressure areas in central and northern Oklahoma. However, scatter plots of understress/excess pore pressure and modeled pore pressure (Figure S2.7) do not show any significant correlation. We also notice that the pore pressure from Mohr circle analysis is much higher than the pore pressure from hydrogeologic models.

The lack of significant correlation between calculated pore pressure from Mohr's circle and pore pressure from hydrogeologic models in Figure S2.7 could be due to the uncertainty of fault geometries during excess pore pressure

calculation. To account for the uncertainties in fault strike and slip from fault mapping, we add to each fault strike and dip a random uncertainty drawn from a normal distribution with standard deviation of 5° and 10°, respectively. Then the understress is calculated using the new fault strike and dip angle. The procedure is repeated for 200 times. The required mean pore pressure with one standard deviation is shown for each fault in Figure S2.8. The uncertainty from the strike and dip alone can cause the required pore pressure to change 2 to 12 MPa. We should notice that the original values of required pore pressure fall within the uncertainty test.

The lack of correlation in Figure S2.7 can also be due to heterogeneity in subsurface hydrogeologic parameters that is not considered in the pore pressure model in Langenbruch et al. (2018). Permeability heterogeneity has been shown to be log-normally distributed in space and therefore certain localized regions may have locally higher pore pressure perturbations than the larger regional pore pressure perturbation. As an example, X. Chen et al. (2018) demonstrated that a highly-permeable fault damage zone can significantly enhance the pore pressure within the fault zone compared to isotropic hydrological structure. These factors likely prevent a strong correlation for individual fault parameters. On the other hand, the stress tensor is derived from events distributed within a larger grid (i.e., much larger than individual dimensions), and represents the averaged effect of regional pore pressure variations, so we cannot see a stronger correlation.

2.5.2 Stress Tensor Heterogeneity

In the analysis above, we use a stress tensor with a spatially uniform amplitude and the local stress orientations for each grid to calculate the understress parameter of the nearby faults. The uniform stress amplitude is taken as the mean value of the calculated absolute principal stress of all grids. To examine the validity of the uniform amplitude measure, we compare the fault stress state based on three different stress field maps: (1) a stress field with uniform stress amplitude (the average stress amplitudes above) and uniform stress orientations with horizontal N85°E σ_1 and vertical σ_2 in a strike-slip faulting regime. (2) a stress field with a uniform stress amplitude and local stress orientations, which is used in the analysis of this study; (3) a stress field with local stress amplitudes and local stress orientations for each grid. Under each scenario, we calculate the understress parameter of each fault and show them in Figure 2.5.

The main difference between scenario (1) and (2) is the spatial variation of principal stress orientations. Results from scenario (1) for some seismogenic faults are consistent with previous studies: e.g., the Prague fault is optimally oriented with understress ~ 0.002, consistent with Marsh and Holland (2016). However, the understress parameter shows differences with local stress orientations considered in scenario (2). Compared to scenario (2), the criticality of the faults in scenario (1) can be either overestimated or underestimated, where 14% of the seismogenic faults have understress changes exceeding 0.1 (Figure 2.5a–b). The difference between two scenarios suggests that the heterogeneity of stress orientations has important implications on the inferred fault stress state.

We further consider the variability of stress amplitude by comparing results from scenario (3) and (2) in Figure 2.5c–d. In contrast to the large difference between scenario (1) and (2), the difference between (3) and (2) is smaller, with only one fault having an understress difference larger than 0.1. This result suggests that the local principal stress orientations have a more significant effect on fault stress state than do the stress amplitudes, and that assuming a uniform stress amplitude does not significantly affect the results.

2.5.3 Effect of the Friction Coefficient

The fault orientation analysis in this study is based on the assumption of constant coefficient of friction of 0.68. Using a uniform friction coefficient for the whole study area, we attribute the fault strength heterogeneity to the variations of fluid pore pressure acting on the fault. The assumption is in part supported by the findings from laboratory experiments on Oklahoma basement rock (Kolawole, Johnston, et al., 2019) and in-situ stress measurement in deep boreholes that the heterogeneity of friction coefficients within different rock types is substantially smaller than the fluid pore pressure heterogeneity (Terakawa et al., 2012). Without further knowledge of the spatial heterogeneity of the friction coefficients, we perform a sensitivity analysis where we vary the coefficient of friction, using constant values of 0.4, 0.6, 0.8, and 1.0 to calculate the stress tensor and understress parameter for seismogenic faults following the same process as for $\mu = 0.68$. The results are shown in Figure 2.6. This experiment demonstrates that the fault understress state is moderately sensitive to the assumed friction coefficient. If we define the optimally oriented faults as those with understress smaller than 0.2, friction coefficients of 0.4, 0.6,

0.8, and 1.0 will identify 72.46%, 76.81%, 75.36%, 71.01% of seismogenic faults as optimally oriented faults, respectively, compared to 78.26% for $\mu = 0.68$. Under a friction coefficient of 0.68, we get the largest percentage of optimally oriented faults. It suggests that we have chosen a value close to the true friction coefficient in the study area. In the future studies, a better knowledge of the spatial distribution of friction coefficient will help further characterize the fault stress state.

2.5.4 Seimogenic Faults, Sedimentary Faults, and the Common Tectonic Control

In this study, seismogenic faults are identified from lineaments of seismicity, so most of the faults are located in the crystalline basement. These faults show different orientations from mapped faults in OGS database (Marsh & Holland, 2016), which is compiled from past literature and data contributed by the oil and gas industry. They are primarily faults in the sedimentary sequences and are referred to as sedimentary faults. Considering that Nemaha fault acts as a pressure boundary in hydrologic modeling, we separate north-central Oklahoma into east and west sections and compare the fault orientations in each section (Figure 2.7). In the west section, the seismogenic faults (Figure 2.7a) show patterns of $[55°, 75°]$ and $[105°, 125°]$, and the sedimentary faults (Figure 2.7b) show a dominant trend of $[45°, 75°]$ and a minor trend of $[0°, 10°]$. For both types of faults, the NE-trending set is more prominent than NW-trending set. It is possibly related to an overwhelming dominance of NE-trending basement-rooted splays (synthetic Reidels) (Liao et al., 2017; Curren & Bird, 2014) distributed along a few large N-S trending basement faults, e.g., the Galena

Township Fault.

In the east section, the seismogenic faults (Figure 2.7c) show excellent correspondence with the observations of basement faults on 3D seismic data in northeast Oklahoma (Kolawole, Johnston, et al., 2019). Both the NE and NW trends are reactivated in the current stress field. The sedimentary faults show dominant trends of NNE to NE, and E-W Figure 2.7d, which are possibly associated with the large basement-rooted NNE faults. The sedimentary faults are poorly oriented in the present-day stress field and do not have earthquakes associated with them currently. Although the NW and NE trends are the most reactivated trends associated with earthquakes, the E-W trend could still pose an important seismic hazard. Localized stress perturbation of the E-W-trending faults in the area results in their seismogenic reactivation as observed within the Jones swarm (Holland, 2013).

In addition to the seismogenic faults and sedimentary faults in north-central Oklahoma, we also include measurements of exposed granite fractures from Mill Creek and Tishomingo in southern Oklahoma. The fractures are mapped at the satellite-scale from Google Earth images with a spatial resolution of 15m. The basement fractures (Figure 2.7e) exhibit similar conjugate patterns as the seismogenic faults and are also consistent with the measurements of Precambrian basement fabrics in Kolawole, Johnston, et al. (2019). Although our comparison of the seismogenic faults with previously mapped sedimentary faults and Precambrian basement fabric likely reveal a common tectonic control, we observe spatial variations in the azimuth of the reactivated fault systems. This variation of reactivated trends may be controlled by both the relative abundance of the basement rooted fault trends (emplaced by past tectonic

events) and the variation of the local stress field across the eastern and western sections of the Oklahoma seismic zone, which might have been influenced by the Nemaha uplift structure.

2.5.5 Faults with M$\geq$ 5.0 Earthquakes

Since 2011, four large earthquakes (M$\geq$ 5.0) have occurred in Oklahoma: the M5.7 Prague earthquake in 2011, and the M5.1 Fairview, M5.8 Pawnee, and M5.0 Cushing earthquakes in 2016. None of these earthquakes occurred along previously mapped faults (e.g., X. Chen et al., 2017; W. L. Yeck et al., 2017). The geometries of the seismogenic faults delineated from seismicity are mostly consistent with the focal mechanism solutions of the mainshocks (Table S1). With an in-situ stress field, we calculate the stress state of the mainshock fault planes from focal mechanism solutions and the corresponding seismogenic faults with the assumption of a constant friction coefficient of $\mu = 0.68$. As shown in Figure 2.8, the faults that hosted the M5.7 Prague, M5.8 Pawnee, and M5.0 Cushing earthquakes have understress smaller than 0.02, suggesting the faults in Prague, Pawnee, and Cushing were critically stressed and failed under a small perturbation of pore pressure. The seismogenic fault in Fairview is the least optimally oriented with an understress parameter of 0.1, and the fault plane of the mainshock has even higher understress of 0.2, which is likely due to the shallower dipping angle of 66°. Goebel et al. (2017) calculated poroelastic stress perturbations in the Fairview area from a group of high-rate injection wells to the northeast. Their results suggest that the poroelastic stress increase at the distance of the Fairview area is about 100kPa, and the fault orientation is about 15° off the optimal orientation that would

receive maximum Coulomb stress change. Their results are consistent with the relatively high understress value obtained here for the Fairview fault. Figure 2.8 also shows that Fairview fault has highest relative shear stress compared to the other three faults, indicating highest frictional strength (Yoshida et al., 2016). The relative frictional strength variations is qualitatively consistent with observations in Wu et al. (2018), where the Fairview fault has highest overall stress drop compared to the other fault zones, similar to the observations for a fluid induced earthquake swarm in Japan (Yoshida et al., 2017).

Gischig (2015) performed numerical modeling to investigate the effect of the fault orientation on rupture propagation, and the results suggest that optimally oriented faults tend to have uncontrolled ruptures that propagate beyond the pressure front, while less optimally oriented faults tend to have ruptures controlled by the extent of the pressurized zones. From this perspective, the understress parameter can provide insight into the fault rupture process, and hence the seismicity distribution for the M5 sequences. To the first order, the Prague, Pawnee, and Cushing sequences, which are on optimally oriented fault planes, are predominantly mainshock-aftershock sequences (Figure S2.9), with large values of skewness of moment release (Zhang & Shearer, 2016), while the Fairview sequence on the least optimally oriented fault is mainly a swarm-type sequence with an extended foreshock sequence leading up to the M5 earthquake resulting in the smallest skewness. Thus, our findings are at least consistent with a hypothesis that the fault criticality influences the temporal evolution of earthquake sequences.

2.5.6 Earthquake Hazard Potential

The maximum magnitude of induced earthquakes during and after injection is essential in evaluating seismic hazard. Several hypotheses have been proposed to understand the maximum magnitude as summarized in Eaton and Igonin (2018). McGarr (2014) proposed that the maximum magnitude can be constrained by the total injection volume and the area of the pressurized zone, which is consistent with the modeling results in Dieterich et al. (2015). In contrast, van der Elst et al. (2016) proposed that the maximum magnitude is related to the magnitude-frequency-distribution of the induced earthquake sequence and related to the b-value and the seismogenic index model proposed by Shapiro et al. (2011). Both models are consistent with the observations. However, it is important to recognize that the fault stress state may have a strong influence on how a rupture grows along the fault. The possibility that quasi-static slip along a pressurized fault grows into dynamic slip beyond the pressurized area has been demonstrated theoretically by Garagash and Germanovich (2012) and verified by a stochastic model in Gischig and Wiemer (2013).

The fault length provides a direct measurement to estimate the maximum magnitude. Both fault orientation and fault length can influence the maximum magnitude of induced earthquakes. In Figure 2.9a, we plot the maximum earthquake magnitude and fault understress state. For optimally oriented faults (understress$<$0.2), the fault has a broader range of magnitudes observed. For non-optimally oriented faults, there are no earthquakes larger than magnitude 4.5. Intermediate to small earthquakes (M$<$ 4.0) occurred on both critically stressed and noncritically stressed faults with a relatively low shear stress,

which could be attributed to increased pore pressure. We map 54 optimally oriented faults and 15 non-optimally oriented faults. It is still possible that large earthquakes have the same probability to occur on non-optimally oriented faults, and the lack of large earthquakes on non-optimally oriented faults is due to its low abundance (Table S2).

Figure 2.9b shows that the maximum magnitude on the faults increases with the fault length. The largest earthquakes (Prague, Pawnee, and Cushing earthquakes) occurred on critically stressed faults and the magnitudes are comparable to the predicted values from the empirical relationship for natural earthquakes (Wells & Coppersmith, 1994), indicating that the largest earthquakes might be controlled by the local stress field. To examine the alternative possibility that large earthquakes control apparent fault size, we plot the seismicity prior and after the mainshock in four M5 clusters (Figure S2.10). The results show that the events before the mainshock already spread over the whole length of the fault in Fairview, Pawnee and Cushing, which suggests that the fault length is not controlled by the mainshock and its aftershocks. Similar observations have been found by Schoenball and Ellsworth (2017a). In Prague, it seems that the fault length is controlled by the mainshock (aftershocks). However, due to the lack of stations at that time, it is also possible that some events before mainshock are missing in the catalog. Based on current data, we cannot draw an unambiguous relationship between understress and maximum magnitude. However, the knowledge of fault stress state can help identify high-risk faults with potential runaway ruptures and large earthquakes (Galis et al., 2017).

2.6 Conclusions

To better characterize the properties and stress state of fault systems in Oklahoma and southern Kansas, we map the fault geometry using high-precision earthquake relocations and generate a high-resolution in-situ stress map using focal mechanism solutions. Our results suggest that:

1. Although the majority of the seismogenic faults (NE and NW trending) are optimally oriented relative to the local stress field, some non-optimally oriented faults are identified.
2. Comparison of the seismogenic faults with sedimentary faults and mapped basement fractures suggests potentially similar tectonic origins for those structures.
3. The orientations of the faults that hosted the largest earthquakes ($M \geq 5.0$) in Oklahoma are quantitatively characterized, and the Prague, Pawnee, and Cushing faults are found to occur on optimally oriented faults, while the Fairview fault is not. For the three large optimally oriented faults, the maximum earthquake magnitudes are comparable to the predictions from the empirical scaling relation for natural earthquakes, and the three sequences are predominately mainshock-aftershock type sequences.

Our study contributes detailed seismogenic fault analysis to the current fault database and provides a more complete picture of the relation among seismogenic fault properties, pore pressure, the local stress field, and rupture process in the region of induced seismicity.

Acknowledgments

This work was supported by the Oklahoma Governor's Emergency Funding for Induced Seismicity, USGS NEHRP external research grant G18AP00022, and graduate student fellowship at the School of Geosciences at the University of Oklahoma. D. Trugman acknowledges institutional support from the Laboratory Directed Research and Development (LDRD) program of Los Alamos National Laboratory under project number 20180700PRD1. We thank the personnel at the Oklahoma Geological Survey (OGS) for sharing focal mechanism and fault data. Focal mechanism data used in this study are included as Supplemental Dataset S1–S2 and available by requesting them from OGS. Fault data are available at `http://www.ou.edu/ogs/data/fault`. We thank the Oklahoma Corporation Commission and the Kansas Corporation Commission for providing wastewater disposal data. We also thank Candace Johnston for data regarding surface fracture orientations in southcentral Oklahoma. Finally, we are grateful for the constructive comments of two anonymous reviewers and the editors, whose suggestions improved the quality and clarity of the manuscript.

2.7 Supplementary Materials

Introduction

The supporting information provides additional information on using the focal mechanism tomography (FMT) technique to calculate absolute stress magnitudes.

Text S1. Here we show the derivation of Equation 2 from the assumptions in Focal Mechanism Tomography (FMT).

Under the four assumptions listed in the manuscript, on the Mohr circle in Figure S2.11, the largest Mohr circle is tangent to the fault strength length line. For the Mohr circle, the center is at ($\frac{\sigma_1+\sigma_3}{2}$,0), and the radius is $\frac{\sigma_1-\sigma_3}{2}$. The fault strength line is $f = \mu(\sigma - P_w) + C$. The distance from a point (x_0,y_0) to a line $ax + by + c = 0$ is defined as:

$$distance(ax + by + c = 0, (x_0, y_0)) = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}} \tag{2.5}$$

Since the largest Mohr circle is tangent to fault strength line, the distance from the center of the largest Mohr circle to fault strength line is equal to the radius, that is:

$$\frac{|\mu(\frac{\sigma_1+\sigma_3}{2} - P_w) + C - 0|}{\sqrt{\mu^2 + 1}} = \frac{\sigma_1 - \sigma_3}{2} \tag{2.6}$$

Then we have,

$$\mu(\frac{\sigma_1 + \sigma_3}{2} - P_w) + C = \frac{\sigma_1 - \sigma_3}{2}\sqrt{\mu^2 + 1} \tag{2.7}$$

By arranging the above equation based on σ_1 and σ_3, we get,

$$\sigma_1(\sqrt{\mu^2+1}-\mu)-\sigma_3(\sqrt{\mu^2+1}+\mu)=2C-2\mu P_w \tag{2.8}$$

Text S2. Here we introduce the method to calculate vertical normal stress from the principal stress tensor following *Allmendinger* (2017). The vertical direction in North-East-Down (NED) system is $\mathbf{n} = (0, 0, 1)^T$. The vector in the coordinate system defined by principal stresses is

$$n_i^{'} = a_{ij} n_j \tag{2.9}$$

where a_{ij} is the transform matrix composed of the orientations (plunge and trend) of principal stresses. The trends and plunges are labeled as $trd_{\sigma i}$ and $plg_{\sigma i}$, where i is the number of the principal stresses. Then:

$$A_{ij} = \begin{pmatrix} \cos(trd_{\sigma 1})\cos(plg_{\sigma 1}) & \sin(trd_{\sigma 1})\cos(plg_{\sigma 1}) & \sin(plg_{\sigma 1}) \\ \cos(trd_{\sigma 2})\cos(plg_{\sigma 2}) & \sin(trd_{\sigma 2})\cos(plg_{\sigma 2}) & \sin(plg_{\sigma 2}) \\ \cos(trd_{\sigma 3})\cos(plg_{\sigma 3}) & \sin(trd_{\sigma 3})\cos(plg_{\sigma 3}) & \sin(plg_{\sigma 3}) \end{pmatrix} \tag{2.10}$$

The vertical stress can be calculated from the principal stress tensor $\sigma' = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix}$ and normal vector $\mathbf{n}^{'}$:

$$\sigma_v = \mathbf{n}^{'T} \sigma' \mathbf{n}' \tag{2.11}$$

References

Alt, R. C., & Zoback, M. D. (2017). In situ stress and active faulting in Oklahoma. *Bulletin of the Seismological Society of America*, *107*(1), 216–228. doi: 10.1785/0120160156

Altmann, J., Müller, B., Müller, T., Heidbach, O., Tingay, M., & Weißhardt, A. (2014). Pore pressure stress coupling in 3d and consequences for reservoir stress states and fault reactivation. *Geothermics*, *52*, 195–205. doi: 10.1016/j.geothermics.2014.01.004

Barbour, A. J., Norbeck, J. H., & Rubinstein, J. L. (2017). The effects of varying injection rates in osage county, oklahoma, on the 2016 m w 5.8 pawnee earthquake. *Seismological Research Letters*, *88*(4), 1040–1053.

Bickford, M., Van Schmus, W., Karlstrom, K., Mueller, P., & Kamenov, G. (2015). Mesoproterozoic-trans-laurentian magmatism: A synthesis of continent-wide age distributions, new sims u–pb ages, zircon saturation temperatures, and hf and nd isotopic compositions. *Precambrian Research*, *265*, 286–312.

Bott, M. H. P. (1959, mar). The mechanics of oblique slip faulting. *Geological Magazine*, *96*(2), 109–117. doi: 10.1017/S0016756800059987

Byerlee, J. (1978). Friction of rocks. *Pure and Applied Geophysics PAGEOPH*, *116*(4-5), 615–626. doi: 10.1007/BF00876528

Cappa, F., Scuderi, M. M., Collettini, C., Guglielmi, Y., & Avouac, J.-P. (2019). Stabilization of fault slip by fluid injection in the laboratory and in situ. *Science advances*, *5*(3), eaau4065. doi: 10.1126/sciadv.aau4065

Chen, C. (2016). *Comprehensive analysis of oklahoma earthquakes: From earthquake monitoring to 3d tomography and relocation* (Doctoral dissertation, University of Oklahoma). Retrieved from `https://shareok.org/handle/11244/45025`

Chen, X., Haffener, J., Goebel, T. H., Meng, X., Peng, Z., & Chang, J. C. (2018). Temporal correlation between seismic moment and injection volume for an induced earthquake sequence in central oklahoma. *Journal of Geophysical Research: Solid Earth*, *123*(4), 3047–3064.

Chen, X., Nakata, N., Pennington, C., Haffener, J., Chang, J. C., He, X., ... Walter, J. I. (2017). The Pawnee earthquake as a result of the interplay among injection, faults and foreshocks. *Scientific Reports*, *7*(1), 1–18. doi: 10.1038/s41598-017-04992-z

Chopra, S., Kurt, M., Folarin, K., & Brett, M. C. (2018). Nemaha strike-slip fault expression on 3-d seismic data in scoop trend. *AAPG Explorer*.

Curren, I. S., & Bird, P. (2014). Formation and suppression of strike–slip fault systems. *Pure and Applied Geophysics*, *171*(11), 2899–2918.

Darold, A., & Holland, A. (2015). Preliminary oklahoma optimal fault orientations. *Oklahoma Geol. Surv. Open File Rep., OF4*.

Deng, K., Liu, Y., & Harrington, R. M. (2016). Poroelastic stress triggering of the December 2013 Crooked Lake, Alberta, induced seismicity sequence. *Geophysical Research Letters*, *43*(16), 8482–8491. doi: 10.1002/2016GL070421

Denison, R. E., Bickford, M., Lidiak, E. G., & Kisvarsanyi, E. (1987). Geology

and geochronology of precambrian rocks in the central interior region of the united states.

Dieterich, J. H., Richards-Dinger, K. B., & Kroll, K. A. (2015). Modeling Injection-Induced Seismicity with the Physics-Based Earthquake Simulator RSQSim. *Seismological Research Letters*, *86*(4), 1102–1109. doi: 10.1785/0220150057

Eaton, D. W., & Igonin, N. (2018). What controls the maximum magnitude of injection-induced earthquakes? *The Leading Edge*, *37*(2), 135–140. doi: 10.1190/tle37020135.1

Ellsworth, W. L. (2013). Injection-induced earthquakes. *Science*, *341*(6142), 1225942. doi: 10.1126/science.1225942

Eyre, T. S., Eaton, D. W., Garagash, D. I., Zecevic, M., Venieri, M., Weir, R., & Lawton, D. C. (2019). The role of aseismic slip in hydraulic fracturing–induced seismicity. *Science advances*, *5*(8), eaav7172. doi: 10.1126/sciadv.aav7172

Galis, M., Ampuero, J. P., Mai, P. M., & Cappa, F. (2017). Induced seismicity provides insight into why earthquake ruptures stop. *Science advances*, *3*(12), eaap7528. doi: 10.1126/sciadv.aap7528

Garagash, D. I., & Germanovich, L. N. (2012). Nucleation and arrest of dynamic slip on a pressurized fault. *Journal of Geophysical Research B: Solid Earth*, *117*(10), 1–27. doi: 10.1029/2012JB009209

Gischig, V. S. (2015). Rupture propagation behavior and the largest possible earthquake induced by fluid injection into deep reservoirs. *Geophysical Research Letters*, *42*(18), 7420–7428. doi: 10.1002/2015GL065072

Gischig, V. S., & Wiemer, S. (2013). A stochastic model for induced seismicity based on non-linear pressure diffusion and irreversible permeability en-

hancement. *Geophysical Journal International*, *194*(2), 1229–1249. doi: 10.1093/gji/ggt164

Goebel, T. H., Weingarten, M., Chen, X., Haffener, J., & Brodsky, E. E. (2017). The 2016 Mw5.1 Fairview, Oklahoma earthquakes: Evidence for long-range poroelastic triggering at >40 km from fluid disposal wells. *Earth and Planetary Science Letters*, *472*, 50–61. doi: 10.1016/j.epsl.2017.05.011

Haffener, J., Chen, X., & Murray, K. (2018). Multiscale analysis of spatiotemporal relationship between injection and seismicity in oklahoma. *Journal of Geophysical Research: Solid Earth*, *123*(10), 8711–8731. doi: 10.1029/2018JB015512

Hardebeck, J. L., & Hauksson, E. (2001). Crustal stress field in southern california and its implications for fault mechanics. *Journal of Geophysical Research: Solid Earth*, *106*(B10), 21859–21882.

Hardebeck, J. L., & Michael, A. J. (2006). Damped regional-scale stress inversions: Methodology and examples for southern california and the coalinga aftershock sequence. *Journal of Geophysical Research: Solid Earth*, *111*(B11). doi: 10.1029/2005JB004144

Hardebeck, J. L., & Shearer, P. M. (2008). *Hash: A fortran program for computing earthquake first-motion focal mechanisms–v1. 2–january 31, 2008.*

Healy, J., Rubey, W., Griggs, D., & Raleigh, C. (1968). The denver earthquakes. *Science*, *161*(3848), 1301–1310. doi: 10.1126/science.161.3848.1301

Holland, A. A. (2013). Optimal fault orientations within oklahoma. *Seismological Research Letters*, *84*(5), 876–890. doi: 10.1785/0220120153

Keranen, K. M., Savage, H. M., Abers, G. A., & Cochran, E. S. (2013). Potentially induced earthquakes in Oklahoma, USA: Links between wastewater

injection and the 2011 Mw 5.7 earthquake sequence. *Geology*, *41*(6), 699–702. doi: 10.1130/G34045.1

Keranen, K. M., Weingarten, M., Abers, G. A., Bekins, B. A., & 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

King Hubbert, M., & Rubey, W. W. (1959). Role of fluid pressure in mechanics of overthrust faulting: I. Mechanics of fluid-filled porous solids and its application to overthrust faulting. *Bulletin of the Geological Society of America*, *70*(2), 115–166. doi: 10.1130/0016-7606(1959)70[115:ROFPIM]2.0.CO;2

Kolawole, F., Carpenter, B., Reches, Z., & Simpson, M. T. (2019). Basement-driven deformation of the sedimentary sequence in north-central oklahoma. In *2019 aapg annual convention and exhibition.*

Kolawole, F., Johnston, C. S., Morgan, C. B., Chang, J. C., Marfurt, K. J., Lockner, D. A., ... Carpenter, B. M. (2019). The susceptibility of Oklahoma's basement to seismic reactivation. *Nature Geoscience*. Retrieved from `https://doi.org/10.1038/s41561-019-0440-5` doi: 10.1038/s41561-019-0440-5

Langenbruch, C., Weingarten, M., & Zoback, M. D. (2018). Physics-based forecasting of man-made earthquake hazards in oklahoma and kansas. *Nature communications*, *9*(1), 3946. doi: 10.1038/s41467-018-06167-4

Levandowski, W., Herrmann, R. B., Briggs, R., Boyd, O., & Gold, R. (2018). An updated stress map of the continental united states reveals heterogeneous intraplate stress. *Nature Geoscience*, 1. doi: https://doi.org/10.1038/s41561-018-0120-x

Levandowski, W., Weingarten, M., & Walsh III, R. (2018). Geomechanical sensitivities of injection-induced earthquakes. *Geophysical Research Letters*, *45*(17), 8958–8965. doi: 10.1029/2018GL077551

Liao, Z., Liu, H., Jiang, Z., Marfurt, K. J., & Reches, Z. (2017). Fault damage zone at subsurface: A case study using 3d seismic attributes and a clay model analog for the anadarko basin, oklahoma. *Interpretation*, *5*(2), T143–T150. doi: 10.1190/INT-2016-0033.1

Marsh, S., & Holland, A. (2016). Comprehensive fault database and interpretive fault map of oklahoma. *Oklahoma Geol. Surv. Open-File Rept.*, 15.

Martínez-Garzón, P., Ben-Zion, Y., Abolfathian, N., Kwiatek, G., & Bohnhoff, M. (2016). A refined methodology for stress inversions of earthquake focal mechanisms. *Journal of Geophysical Research: Solid Earth*, *121*(12), 8666-8687. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JB013493` doi: 10.1002/2016JB013493

Martínez-Garzón, P., Bohnhoff, M., Kwiatek, G., & Dresen, G. (2013). Stress tensor changes related to fluid injection at the geysers geothermal field, california. *Geophysical Research Letters*, *40*(11), 2596–2601. doi: 10.1002/grl.50438,2013

Martínez-Garzón, P., Kwiatek, G., Bohnhoff, M., & Dresen, G. (2016). Impact of fluid injection on fracture reactivation at the geysers geothermal field. *Journal of Geophysical Research: Solid Earth*, *121*(10), 7432–7449. doi: 10.1002/2016JB013137

Martínez-Garzón, P., Kwiatek, G., Ickrath, M., & Bohnhoff, M. (2014). MSATSI: A MATLAB Package for Stress Inversion Combining Solid Classic Methodology, a New Simplified User-Handling, and a Visualization Tool. *Seismological Research Letters*, *85*(4), 896–904. doi: 10.1785/

0220130189

Martínez-Garzón, P., Vavryčuk, V., Kwiatek, G., & Bohnhoff, M. (2016). Sensitivity of stress inversion of focal mechanisms to pore pressure changes. *Geophysical Research Letters*, *43*(16), 8441–8450. doi: 10.1002/2016GL070145

McGarr, A. (2014). Maximum magnitude earthquakes induced by fluid injection. *Journal of Geophysical Research: Solid Earth*, *119*(2), 1008–1019.

Michael, A. J. (1984). Determination of stress from slip data: faults and folds. *Journal of Geophysical Research*, *89*(B13), 11517–11526. doi: 10.1029/JB089iB13p11517

Northcutt, R. A., & Campbell, J. A. (1996). Geologic provinces of oklahoma. , 128–134.

Qi, W. (2016). *Stress Analysis of Recent Earthquakes in Oklahoma* (Master's thesis, University of Oklahoma). Retrieved from `http://hdl.handle.net/11244/44223`

Quinones, L. A., DeShon, H. R., Magnani, M. B., & Frohlich, C. (2018). Stress orientations in the fort worth basin, texas, determined from earthquake focal mechanismsstress orientations in the fort worth basin. *Bulletin of the Seismological Society of America*, *108*(3A), 1124. Retrieved from `http://dx.doi.org/10.1785/0120170337` doi: 10.1785/0120170337

Raleigh, C. B., Healy, J. H., & Bredehoeft, J. D. (1976). An experiment in earthquake control at Rangely, Colorado. *Science*, *191*(4233), 1230–1237. doi: 10.1126/science.191.4233.1230

Rubinstein, J. L., Ellsworth, W. L., & Dougherty, S. L. (2018). The 2013–2016 induced earthquakes in harper and sumner counties, southern kansas. *Bulletin of the Seismological Society of America*, *108*(2), 674–689. doi:

10.1785/0120170209

Schoenball, M., & Ellsworth, W. L. (2017a). A systematic assessment of the spatiotemporal evolution of fault activation through induced seismicity in oklahoma and southern kansas. *Journal of Geophysical Research: Solid Earth*, *122*(12). doi: 10.1002/2017JB014850

Schoenball, M., & Ellsworth, W. L. (2017b). Waveform-Relocated Earthquake Catalog for Oklahoma and Southern Kansas Illuminates the Regional Fault Network. *Seismological Research Letters*. doi: 10.1785/0220170083

Segall, P., & Lu, S. (2015). Injection-induced seismicity: Poroelastic and earthquake nucleation effects. *Journal of Geophysical Research: Solid Earth*, *120*(7), 5082–5103. doi: 10.1002/2015JB012060

Shapiro, S. A., Krüger, O. S., Dinske, C., & Langenbruch, C. (2011). Magnitudes of induced earthquakes and geometric scales of fluid-stimulated rock volumes. *Geophysics*, *76*(6), WC55–WC63. doi: 10.1190/geo2010-0349.1

Skoumal, R. J., Kaven, J. O., & Walter, J. I. (2019). Characterizing seismogenic fault structures in oklahoma using a relocated template-matched catalog. *Seismological Research Letters*. doi: 10.1785/0220190045

Snee, J.-E. L., & Zoback, M. D. (2016). State of stress in texas: Implications for induced seismicity. *Geophysical Research Letters*, *43*(19), 10–208.

Terakawa, T., Miller, S. A., & Deichmann, N. (2012). High fluid pressure and triggered earthquakes in the enhanced geothermal system in Basel, Switzerland. *Journal of Geophysical Research: Solid Earth*, *117*(7), 1–15. doi: 10.1029/2011JB008980

Terakawa, T., Zoporowski, A., Galvan, B., & Miller, S. A. (2010). High-pressure fluid at hypocentral depths in the L'Aquila region inferred from earthquake focal mechanisms. *Geology*, *38*(11), 995–998. doi:

10.1130/G31457.1

van der Elst, N. J., Page, M. T., Weiser, D. A., Goebel, T. H., & Hosseini, S. M. (2016). Induced earthquake magnitudes are as large as (statistically) expected. *Journal of Geophysical Research: Solid Earth*, *121*(6), 4575–4590. doi: 10.1002/2016JB012818

Vavryčuk, V. (2014). Iterative joint inversion for stress and fault orientations from focal mechanisms. *Geophysical Journal International*, *199*(1), 69–77. doi: 10.1093/gji/ggu224

Vidale, J. E., & Shearer, P. M. (2006). A survey of 71 earthquake bursts across southern California: Exploring the role of pore fluid pressure fluctuations and aseismic slip as drivers. *Journal of Geophysical Research: Solid Earth*, *111*(5), 1–12. doi: 10.1029/2005JB004034

Waldhauser, F. (2001). hypodd–a program to compute double-difference hypocenter locations (hypodd version 1.0-03/2001). *US Geol. Surv. Open File Rep., 01*, *113*.

Wallace, R. E. (1951). Geometry of shearing stress and relation to faulting. *The Journal of Geology*, *59*(2), 118–130. doi: 10.1086/625831

Walsh, F. R., & Zoback, M. D. (2016). Probabilistic assessment of potential fault slip related to injectioninduced earthquakes: Application to north-central Oklahoma, USA. *Geology*, *44*(12), 991–994. doi: 10.1130/G38275.1

Walter, J. I., P. Ogwari, A. T., F. Ferrer, I. W., Chang, J. C., Darold, A. P., & Holland, A. A. (2019). The oklahoma geological survey statewide seismic network. *Seismol. Res. Lett.*.

Wells, D. L., & Coppersmith, K. J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the seismological Society of America*, *84*(4),

974–1002.

Whitmeyer, S. J., & Karlstrom, K. E. (2007). Tectonic model for the proterozoic growth of north america. *Geosphere*, *3*(4), 220–259.

Wu, Q., Chapman, M., & Chen, X. (2018). Stress-drop variations of induced earthquakes in oklahoma. *Bulletin of the Seismological Society of America*, *108*(3A), 1107–1123. doi: 10.1785/0120170335

Yeck, W., Weingarten, M., Benz, H. M., McNamara, D. E., Bergman, E., Herrmann, R., ... Earle, P. (2016). Far-field pressurization likely caused one of the largest injection induced earthquakes by reactivating a large preexisting basement fault structure. *Geophysical Research Letters*, *43*(19). doi: 10.1002/2016GL070861

Yeck, W. L., Hayes, G. P., McNamara, D. E., Rubinstein, J. L., Barnhart, W. D., Earle, P. S., & Benz, H. M. (2017). Oklahoma experiences largest earthquake during ongoing regional wastewater injection hazard mitigation efforts. *Geophysical Research Letters*, *44*(2), 711–717. doi: 10.1002/2016GL071685

Yoshida, K., Hasegawa, A., & Yoshida, T. (2016). Temporal variation of frictional strength in an earthquake swarm in ne japan caused by fluid migration. *Journal of Geophysical Research: Solid Earth*, *121*(8), 5953–5965. doi: 10.1002/2016JB013022

Yoshida, K., Saito, T., Urata, Y., Asano, Y., & Hasegawa, A. (2017). Temporal changes in stress drop, frictional strength, and earthquake size distribution in the 2011 yamagata-fukushima, ne japan, earthquake swarm, caused by fluid migration. *Journal of Geophysical Research: Solid Earth*, *122*(12), 10–379. doi: 10.1002/2017JB014334

Zhang, Q., & Shearer, P. M. (2016). A new method to identify earthquake

swarms applied to seismicity near the san jacinto fault, california. *Geophysical Journal International*, *205*(2), 995–1005. doi: 10.1093/gji/ggw073

Zoback, M. L. (1992). First-and second-order patterns of stress in the lithosphere: The world stress map project. *Journal of Geophysical Research: Solid Earth*, *97*(B8), 11703–11728. doi: 10.1029/92JB00132

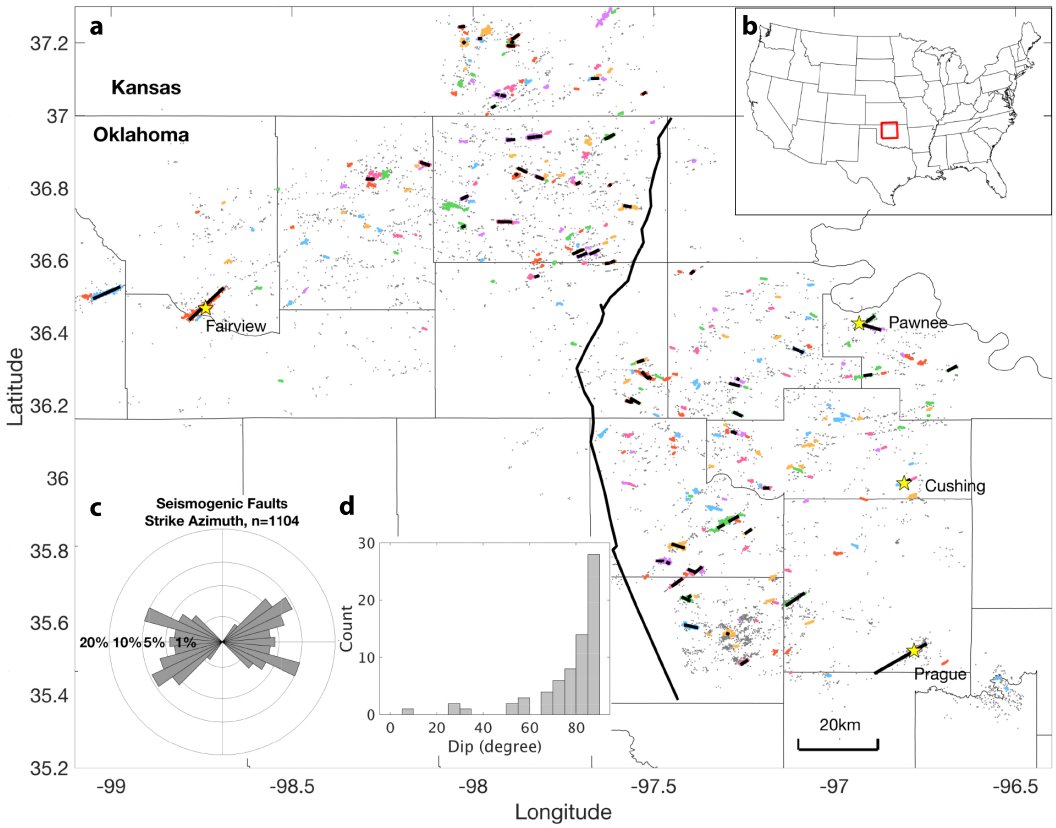

Figure 2.1: Mapped faults in this study based on earthquake relocations from C. Chen (2016) and Schoenball and Ellsworth (2017b). (a) Earthquake clusters (colored dots) with 10 and more events. Short, black lines show the faults with planarity larger than 0.8 mapped from 30 and more events. Yellow stars show the location of four M≥5 earthquakes in Oklahoma. Thin, black lines are county boundaries in Oklahoma. The long, thick black line is the Nemaha fault from OGS fault database (Marsh & Holland, 2016). (b) The inset map shows the location of the study area. (c) Histogram of strike of faults in (a). (d) Histogram of dip of faults in (a).

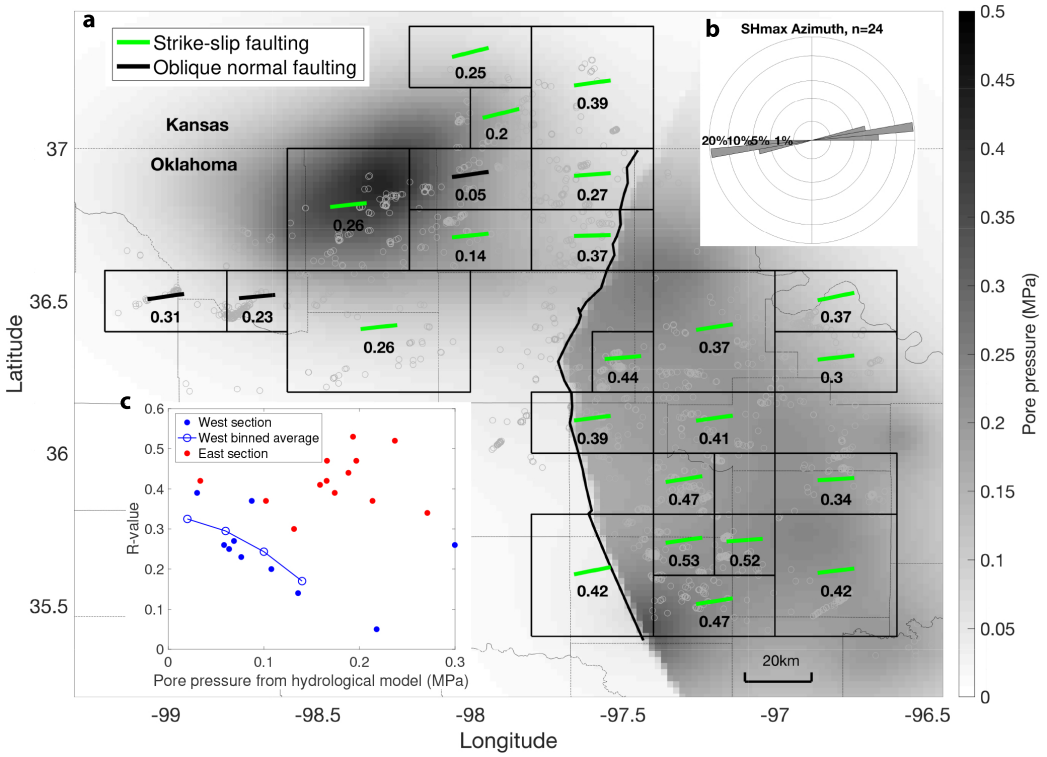

Figure 2.2: Stress inversion results for Oklahoma and southern Kansas. (a) The bars show the orientation of the maximum horizontal stress (σ_{Hmax}) axis. Green: strike-slip faulting; black: oblique normal faulting. Faulting regime is assigned according to Zoback (1992). The number in the grid shows the corresponding R value. Gray circles are focal mechanisms used in the inversion. Thin, gray lines are county boundaries in Oklahoma. The long, thick black line is the Nemaha fault. (b) Rose diagram of σ_{Hmax} orientation. (c) Cross plot of R value and calculated pore pressure from Langenbruch et al. (2018). The background shows the calculated pore pressure by November 2016 from hydrogeologic models (Langenbruch et al., 2018).

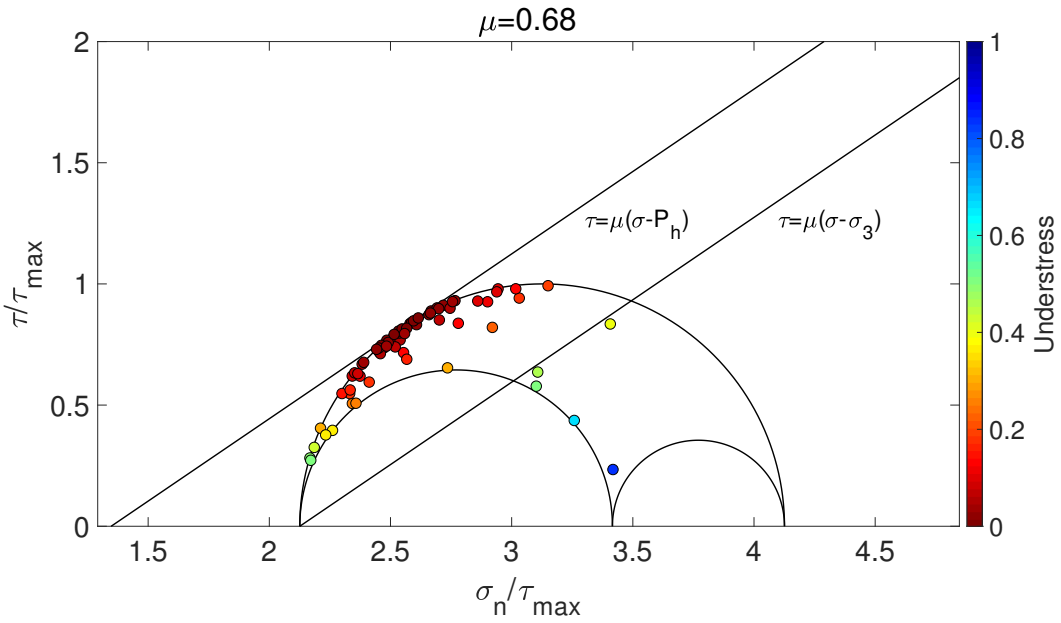

Figure 2.3: The stress state of 69 seismogenic faults in a 3D Mohr diagram. The three semicircles represent the stress tensor, and the two straight lines represent the fault strength under hydrostatic fluid pressure and lithostatic pressure under friction coefficient of 0.68. Each circle, colored by understress value, represents the shear and normal stress on a single fault.

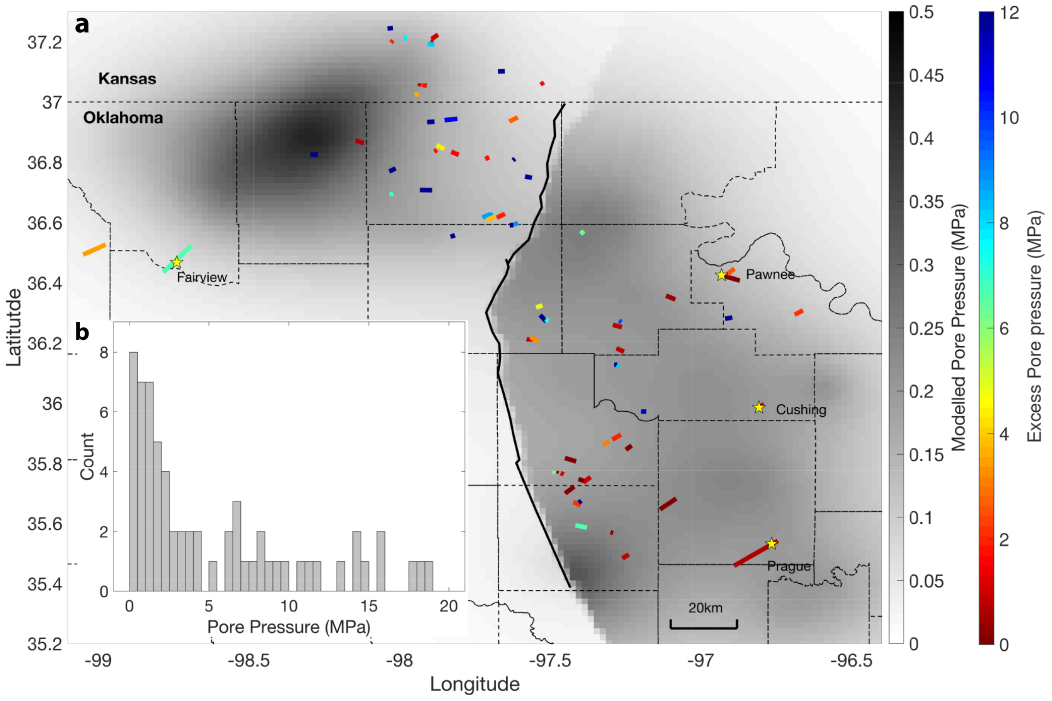

Figure 2.4: (a) Seismogenic fault map colored by excess pore pressure. Yellow stars show the location of four M≥5 earthquakes in Oklahoma. The dashed lines are county boundaries. The long, thick black line is the Nemaha fault. The background shows the calculated pore pressure by November, 2016 from Langenbruch et al. (2018). (b) Histogram of excess pore pressure on each fault.

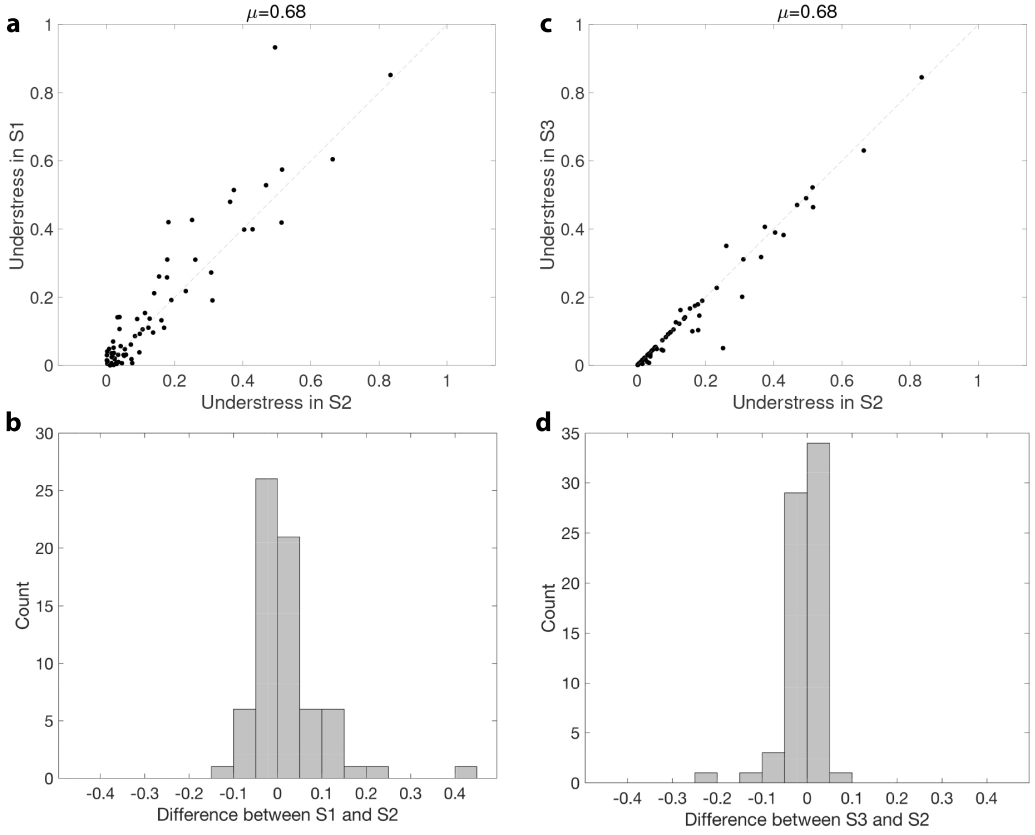

Figure 2.5: Fault understress values from each stress field scenario. (a) Crossplot of understress values between scenario S1 and S2. (b) Histogram of understress difference between scenario S2 and S1. (c) Crossplot of understress values between scenario S3 and S2. (d) Histogram of understress difference between scenario S3 and S2.

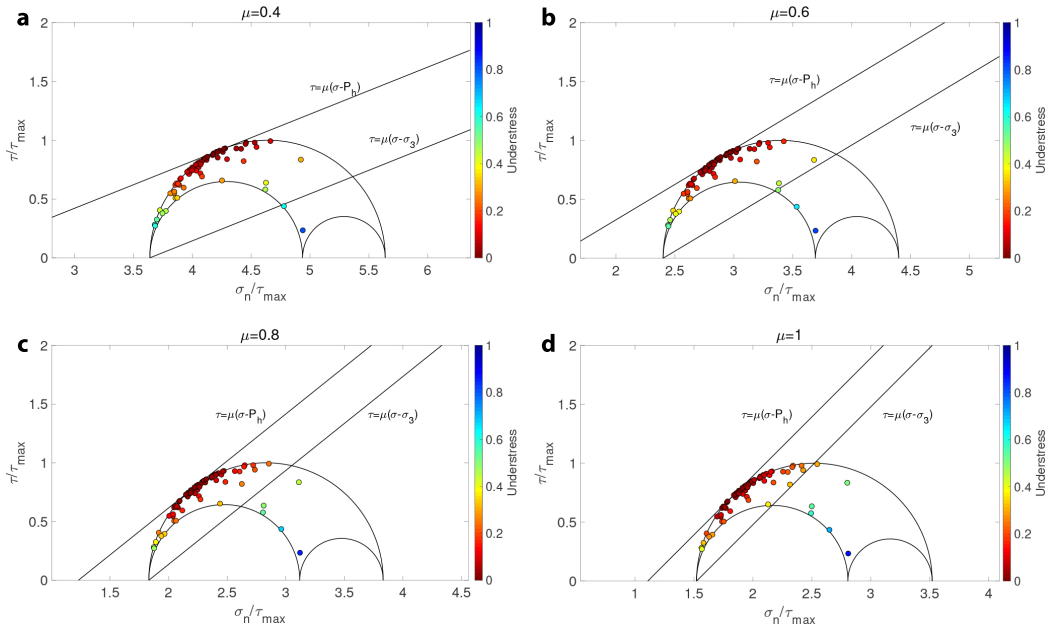

Figure 2.6: Dependence of the fault stress state on friction coefficients of 0.4, 0.6, 0.8, and 1.0. The plot schemes are the same as in Figure 2.3.

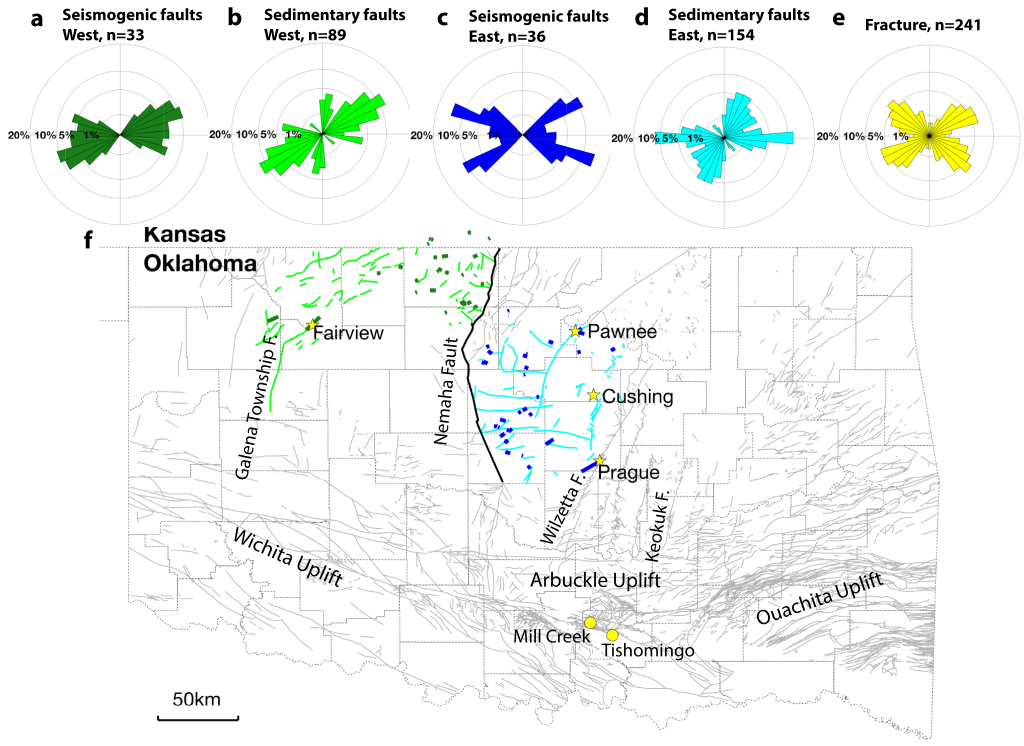

Figure 2.7: Strike comparisons of different categories of faults. (a) Seismogenic faults in the western pressurized section. (b) Mapped sedimentary faults in the western section. (c) Seismogenic faults in the eastern pressurized section. (d) Mapped sedimentary faults in the eastern section. (e) Mapped fractures in Mill Creek and Tishomingo, Oklahoma. The mapped sedimentary faults are from Marsh and Holland (2016). (f) The fault map with the same color scheme as the rose diagrams. The western and eastern pressurized sections are separated by the Nemaha Fault. Yellow dots denote the locations of exposed fractures. Some long N-S trending faults from (Marsh & Holland, 2016) are labeled, including the Nemaha Fault, Wilzetta Fault, Keokuk Fault, and the Galena Township Fault. The figure also labels the geological province in the south of Oklahoma (Northcutt & Campbell, 1996).

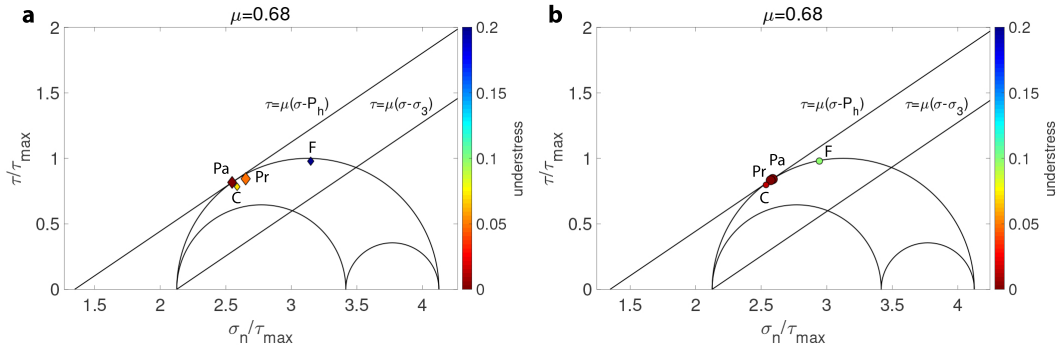

Figure 2.8: Understress for (a) the focal mechanism solutions of M≥5.0 mainshocks and (b) the hosting seismogenic faults as listed in Table S1. Diamonds: focal mechanism solutions. Circles: seismogenic faults. Both symbols are colored by understress values and scaled with the magnitude of the mainshock. Pr: Prague; Pa: Pawnee; C: Cushing; F: Fairview.

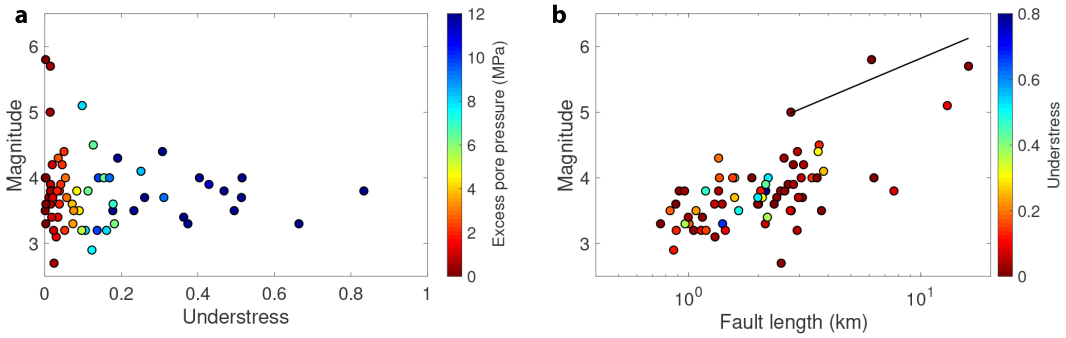

Figure 2.9: (a) Scatter plot of the maximum earthquake magnitude and understress value on each fault. The dots are colored by excess pore pressure. (b) Scatter plot of maximum earthquake magnitude and fault length. The dots are colored by understress value. The black line is the empirical relation between magnitude and rupture length from natural earthquakes: M= 4.33 + 1.49 log(RLD) (M>4.5), RLD–subsurface rupture length (km) (Wells & Coppersmith, 1994).

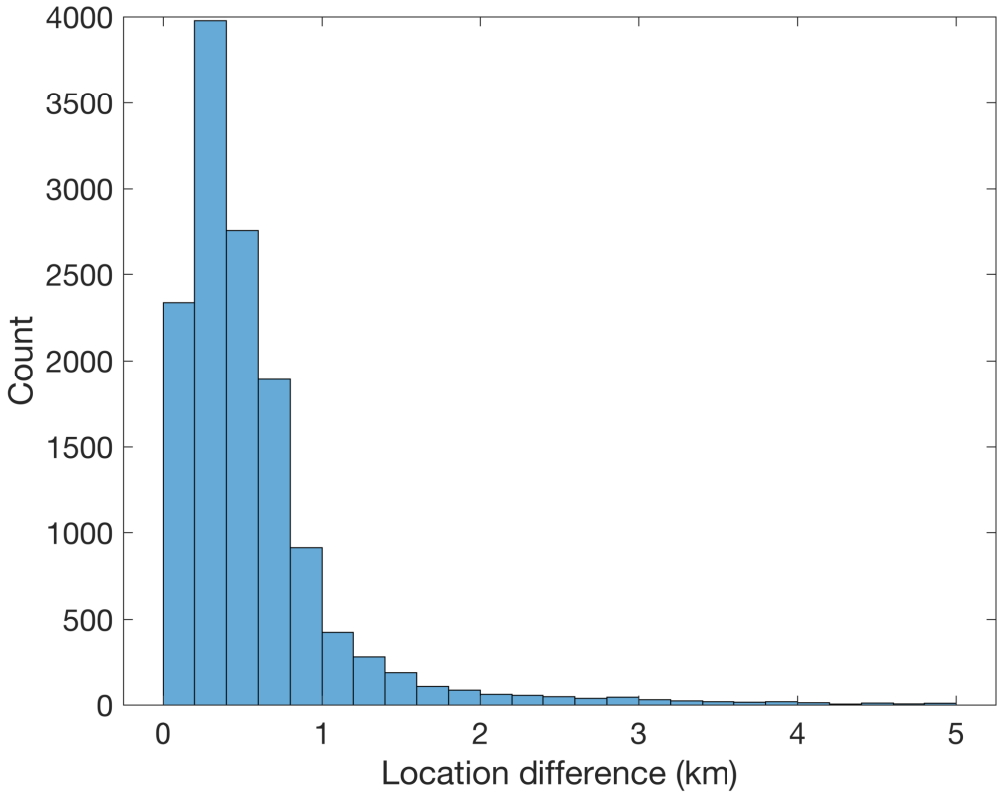

Figure S2.1: Histogram of location difference for common events in the catalog from C. Chen (2016) and Schoenball and Ellsworth (2017b).

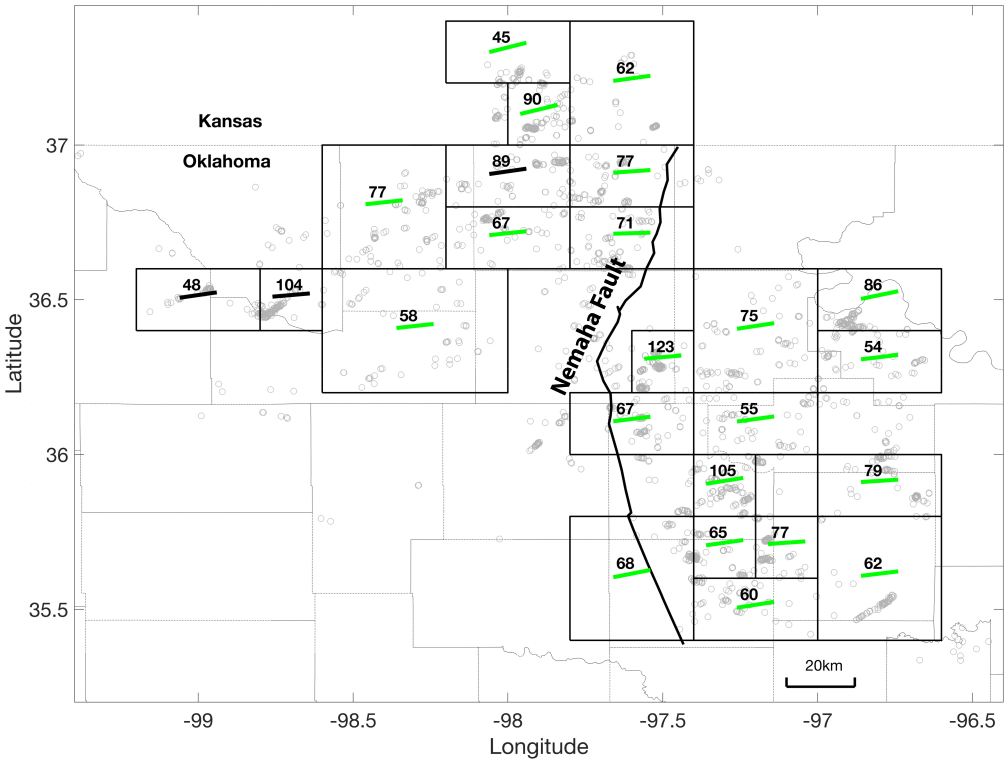

Figure S2.2: Grids and focal mechanism solutions used in stress inversion. The number of focal mechanisms available is shown in each grid. Gray circles are focal mechanisms used in the inversion. The bars show the orientation of the maximum horizontal stress (σ_{Hmax}) axis from inversion. Green: strike-slip faulting; black: oblique normal faulting. Faulting regime is assigned according to Zoback (1992). Thin, gray lines are county boundaries in Oklahoma. The long, thick black line is the Nemaha fault.

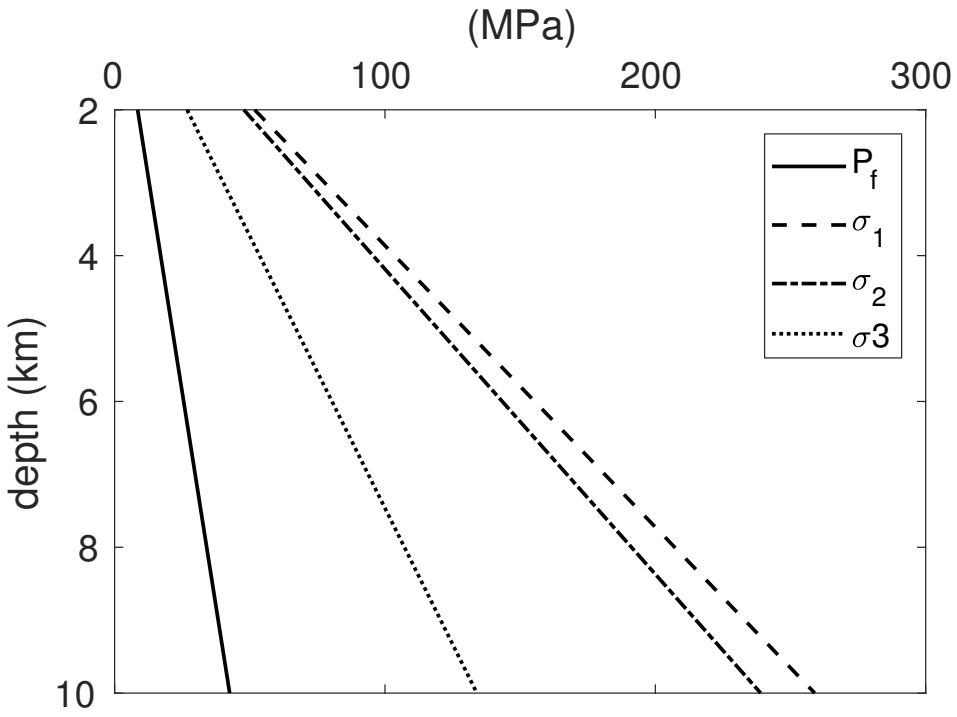

Figure S2.3: An example of principal stress and excess pore pressure (required pore pressure increase for fault failure) changes with depth.

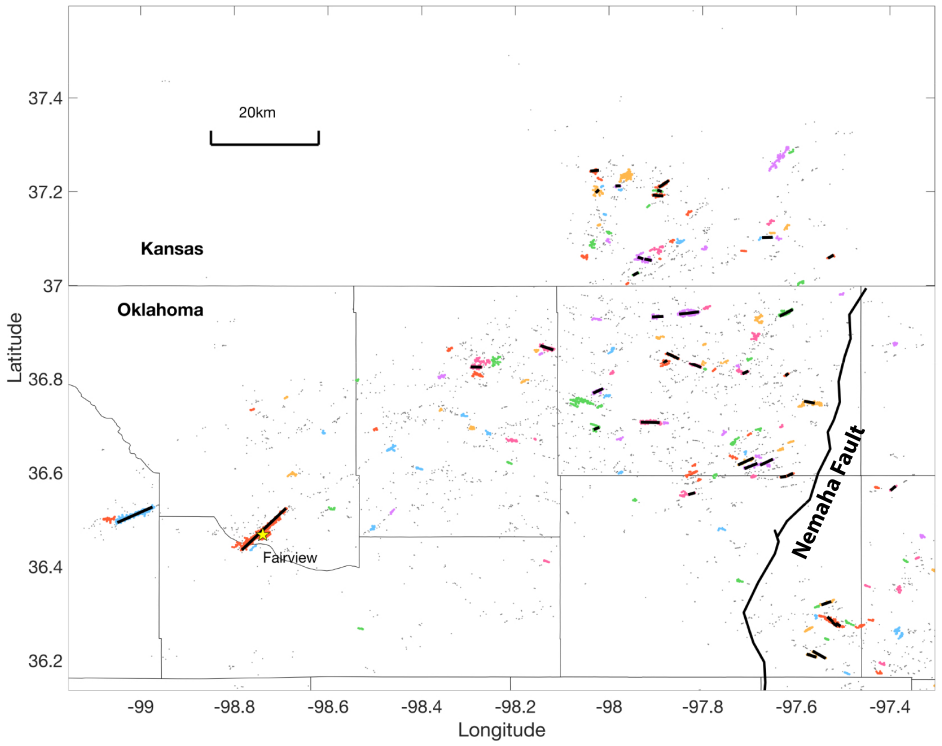

Figure S2.4: A close-up view (west of north-central Oklahoma) of the clusters.

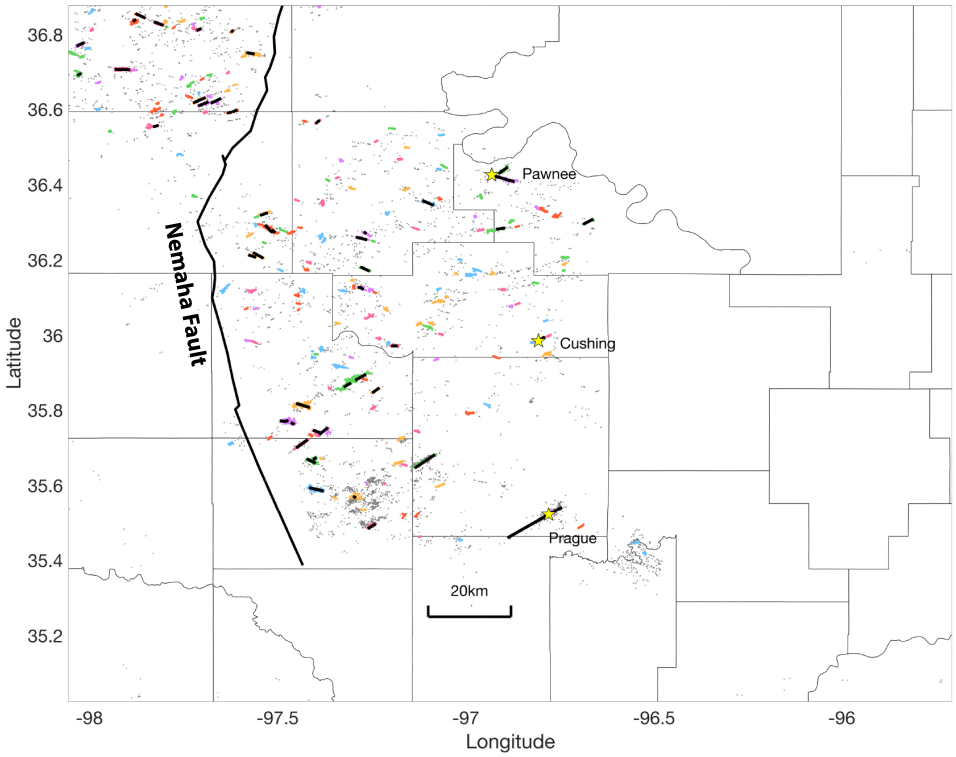

Figure S2.5: A close-up view (east of north-central Oklahoma) of the clusters.

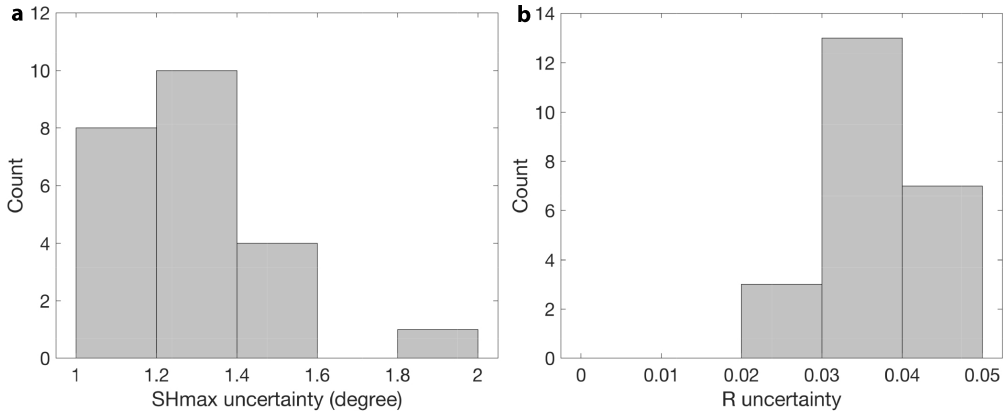

Figure S2.6: The uncertainty of (a) σ_{Hmax} orientation and (b) R value estimated from bootstrap resamplings in stress inversion. Small uncertainties of σ_{Hmax} and R value suggest that the heterogeneity of the stress field is well constrained using the high quality focal mechanism solutions.

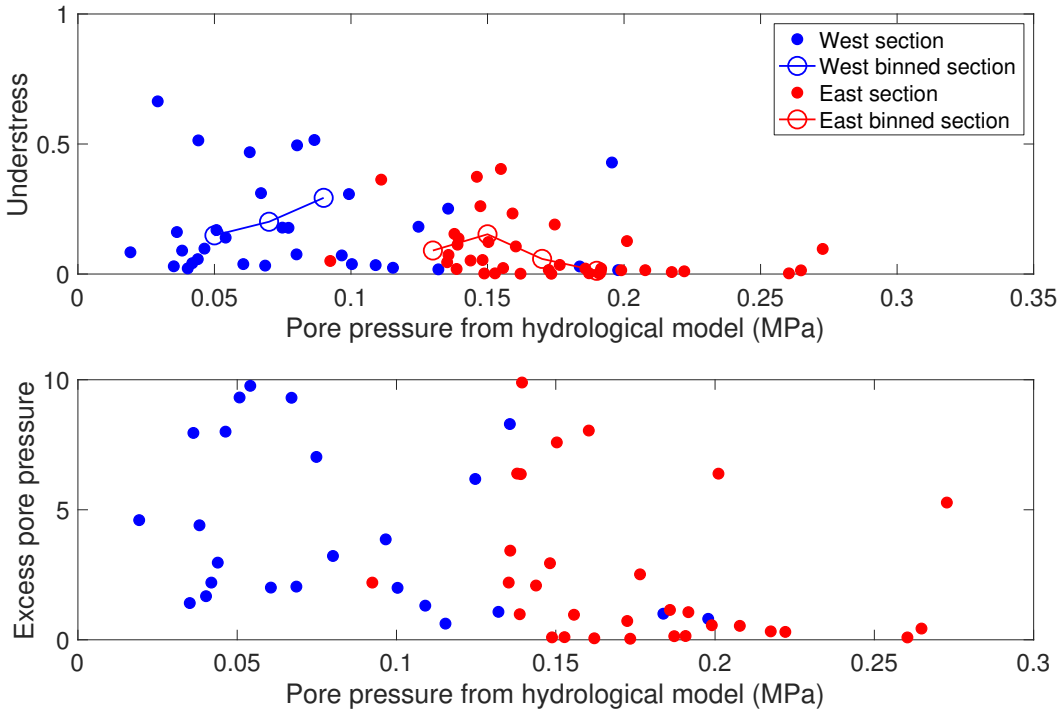

Figure S2.7: (a) Cross plot of understress values and averaged pore pressure on each fault. (b) Cross plot of excess pressure values and averaged pore pressure on each fault. East and west sections are separated by the Nemaha fault. We bin the pore pressure with a bin size of 0.02MPa. The open circles in (a) show the average values of understress which fall in the pore pressure bin of 0.02MPa. The pore pressure was calculated from hydrological model by Langenbruch et al. (2018).

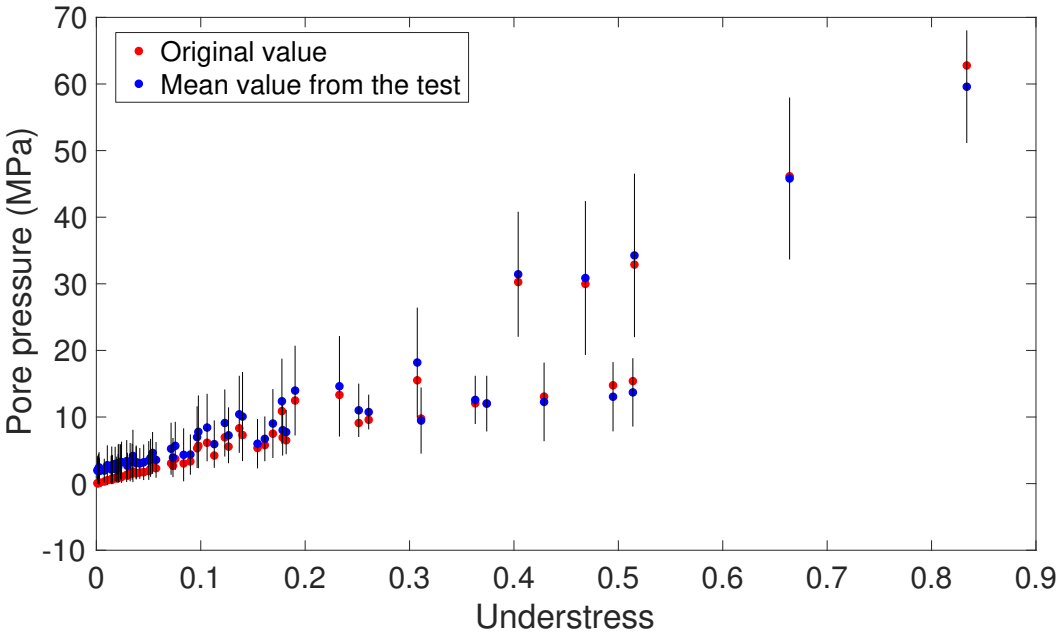

Figure S2.8: Required pore pressure calculated by adding uncertainties to fault strike and slip. The red points are the original pore pressure. The blue points are the mean pore pressure by varying fault strike and dip 200 times. The black lines show the range of one standard deviation.

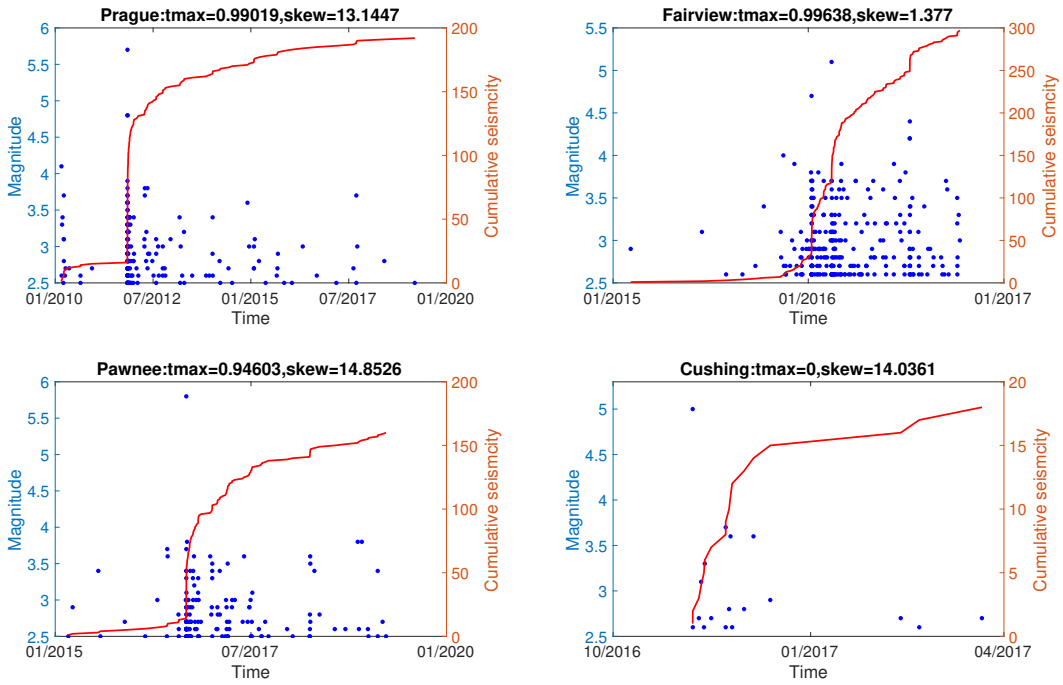

Figure S2.9: The magnitude-time plot (blue dots) and cumulative seismicity rate (red lines) of earthquake sequences in (a) Prague, (b) Fairview, (3) Pawnee, and (4) Cushing, Oklahoma. The raw catalog from the Oklahoma Geological Survey is used in the plot. For Fairview and Cushing, the M5 sequences are selected based on temporal clustering of the whole sequence.

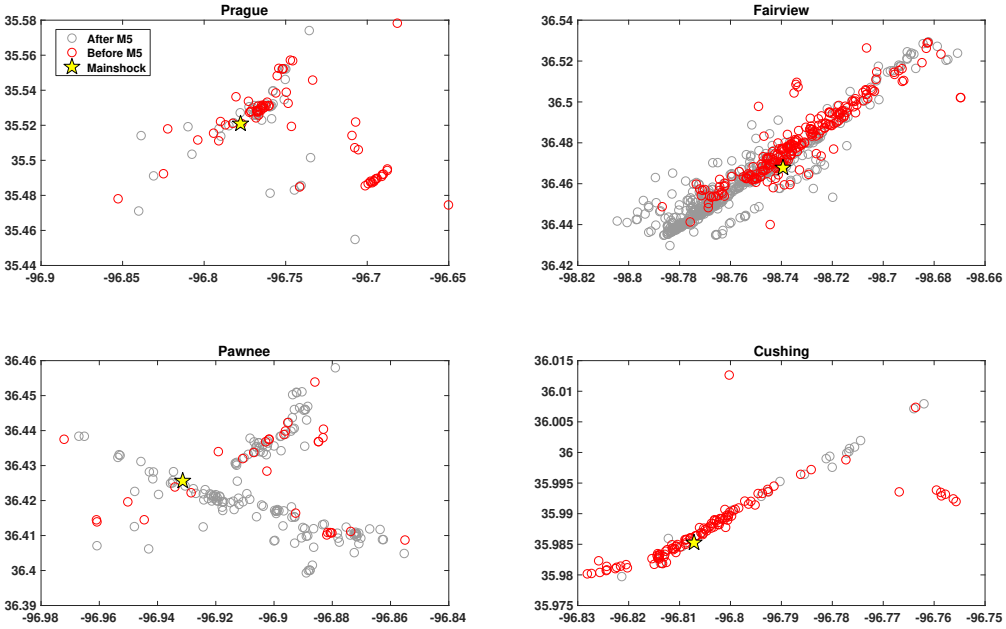

Figure S2.10: Seismicity distribution in the four M5 clusters. The yellow stars are the mainshock. Events before mainshock are in red, after in gray. The event locations are from Schoenball and Ellsworth (2017b) catalog.

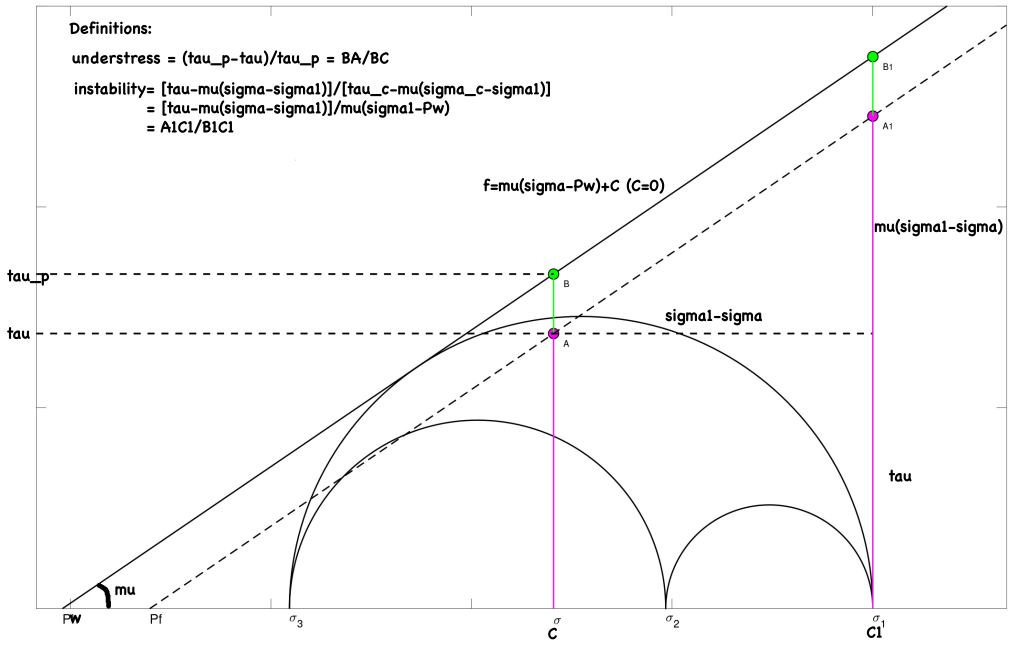

Figure S2.11: Definition of understress (Gischig, 2015) and instability (Vavryčuk, 2014). The largest Mohr circle is tangent to the fault strength line under hydrostatic pore pressure based on the assumptions listed in the manuscript. By definition, understress is calculated as $|BA|/|BC|$, and instability is calculated as $|A_1C_1|/|B_1C_1|$.

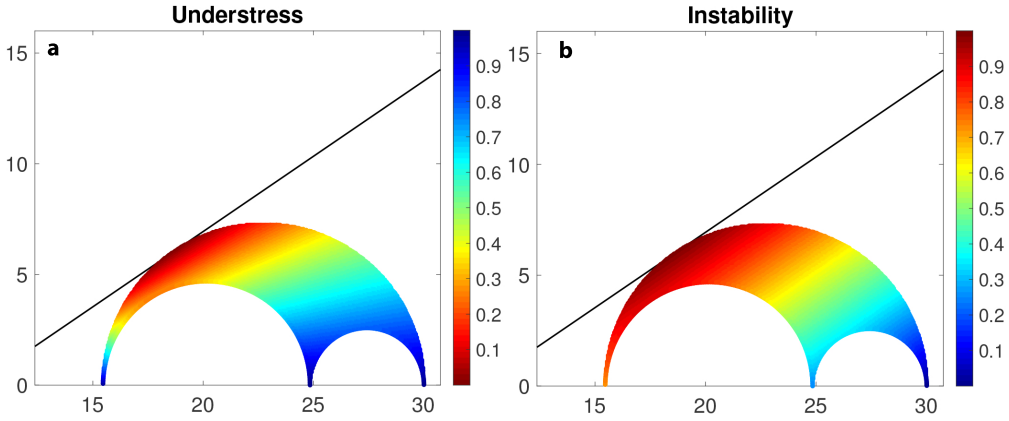

Figure S2.12: Stress state of faults of various orientations represented by (a) understress, and (b) instability. Both parameters can reflect the criticality of the faults.

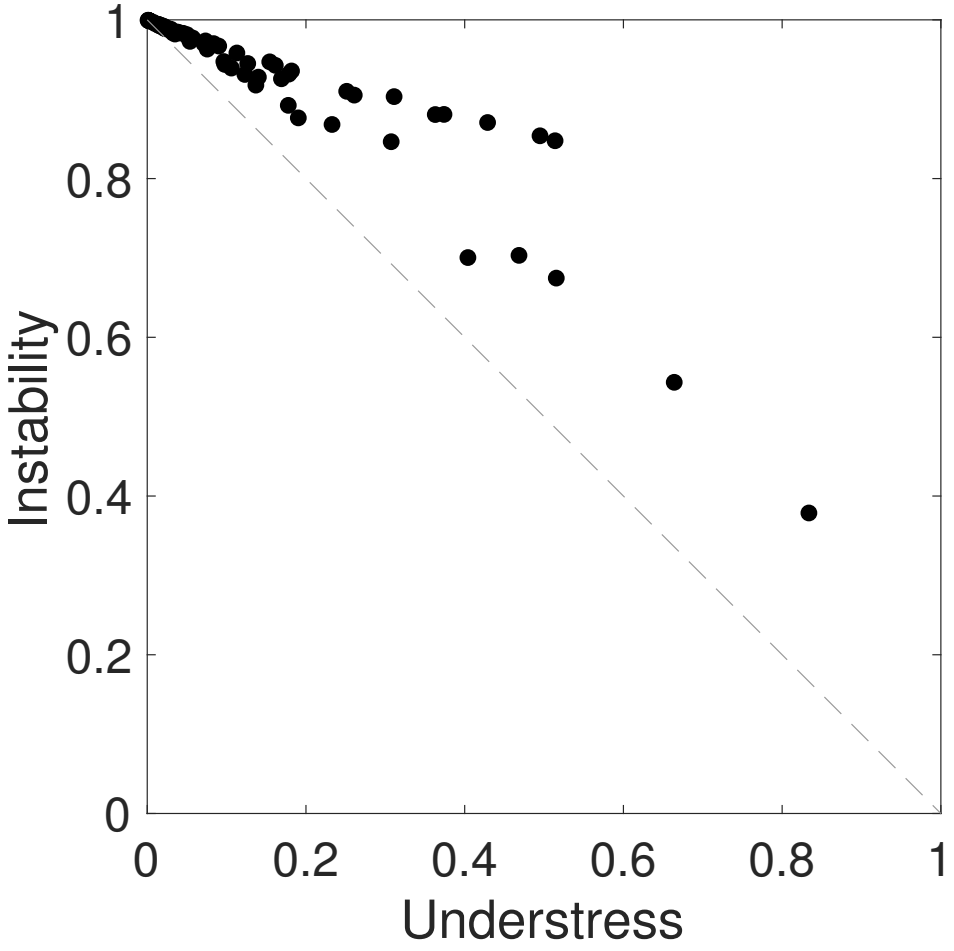

Figure S2.13: Cross plot of understress and instability for 69 seismogenic faults. The bifurcation at high instability values corresponds to the area between the largest and second largest Mohr circles in Figure S2.12.

Table S2.1: Focal Mechanisms of $M \geq 5.0$ Earthquakes and Mapped Fault Geometry.

Area	Mag	MS strike(°)	MS dip(°)	Fault strike(°)	Fault dip(°)
Prague	5.7	236	85	235	86
Fairview	5.1	46	66	46	84
Pawnee	5.8	289	72	-73	89
Cushing	5.0	59	78	59	89

Table S2.2: Statistics of large earthquakes on optimally and non-optimally oriented faults.

Fault Groups	Num	PCT	EQs ($M \geq 5.$)	EQs ($M \geq 4.5$)	EQs ($M \geq 4.$)
Optimally Oriented	54	78%	4	5	17/54=0.31
Misoriented	15	22%	0	0	4/15=0.27

Chapter 3

Diverse Stress State Evolution of Induced Earthquake Sequences in Oklahoma Revealed through High-resolution Focal Mechanisms

3.1 Introduction

The seismicity rate in Oklahoma has experienced significant variations during the last decades, which has been mostly attributed to wastewater disposal (e.g., Ellsworth, 2013; Keranen et al., 2014; Yeck et al., 2016). With

the improved monitoring network in Oklahoma, high-resolution seismicity relocations have revealed previously unknown fault structures and rupture processes. For example, Schoenball and Ellsworth (2017a) and Skoumal et al. (2019) systematically mapped faults through relocated seismicity in Oklahoma; at local scale, the fault structure was mapped using seismicity in Prauge (Keranen et al., 2013), Fairview (Yeck et al., 2016), Pawnee (X. Chen et al., 2017), Guthrie-Langston (Schoenball et al., 2018), and Woodward (Qin et al., 2019).

Earthquake focal mechanisms also provide critical information about the stress field, fault structure, and deformation process. Constraining the focal mechanism of small earthquakes remains difficult because of the low signal-to-noise ratio and a lack of azimuthal coverage in areas. New techniques have been developed to improve focal mechanism determinations and uncertainty characterization. Vavryčuk (2015) inverted for a composite moment tensor using joint inversion of multiple earthquake data. Pugh et al. (2016) and De Matteis et al. (2016) inverted moment tensor from various inputs using Bayesian frameworks. Spatially concentrated earthquakes, such as earthquake swarms and mainshock-aftershock sequences, often show similar mechanisms. Shelly et al. (2016) developed a new strategy for earthquake focal mechanism using waveform-correlation-derived relative polarities and cluster analysis. The strategy clustered events with similar patterns of polarities and applied focal-mechanism inversion to the grouped polarity data. Y. Chen and Huang (2019) developed a moment-tensor joint inversion method for single-borehole geophone array, where the events were clustered based on waveform similarities and radiation patterns, a solution was jointly inverted for the clustered events and then the moment tensors of individual events in each cluster were inverted

with a limited searching range based on the joint inversion results. With the increasing application of machine learning methods in seismology, Ross et al. (2018) trained a convolutional neural network to pick first-motion polarities of P waveform and demonstrated that the networks could pick polarities more accurately than seismic analysts.

In this study, we invert the focal mechanisms and compute the stress state of individual earthquakes in four selected sequences Cushing, Guthrie, Woodward, and Fairview. Previous observations have shown the complexity of the four sequences. In Cushing, where the critical oil pipeline and storage facility are located, the M5.0 earthquake caused structural damage in the nearby city of Cushing (McGarr & Barbour, 2017). It is of great importance to better understand the stress state and rupture process of the M5.0 sequence. In Guthrie, X. Chen et al. (2018) associated the seismicity to the nearest disposal well and found the maximum seismic moment is about an order of magnitude lower than expected from the theory McGarr and Barbour (2017). Wu et al. (2019) found that the M_w 4.0 in Guthrie had a complex rupture process. The findings suggest complex local fault conditions and their influence on seismicity. In Woodward, Qin et al. (2018) mapped a complex fault system of flower structures rather than the prevalent vertical strike-slip faults in central and northern Oklahoma. In Fairview, a M5.1 earthquake occurred in February, 2016 and the sequence kept migrating to southwest, showing a growing length of the main fault, which could cause continuing seismic hazard. Also in Fairview, the sequence was possibly influenced by both pore pressure (McGarr & Barbour, 2017) and poroelastic stress change from more distant wells (Goebel et al., 2017).

In this paper, we use two different methods to obtain the first-motion polarities of small earthquakes and then invert the focal mechanisms for each individual event. Based on the relocation and focal mechanism solutions, we interpret the fault structures and compute the fault stress state. The spatiotemporal analysis of the earthquake relocations and fault stress state shows the detailed fault rupture process. Combined with a modeled pore pressure field, the results are further used to evaluate earthquake hazards.

3.2 Data

We use the earthquake catalog from the Oklahoma Geological Survey (OGS) (Walter et al., 2020). In the four sequences in Cushing, Guthrie, Woodward, Fairview, there are 681, 1073, 1404, and 3460 cataloged events with magnitude range from 0.1 to 5.1 (Figure 3.1) in the selected time period from January, 2009 to August, 2018. The catalog covers the initiation and the following propagation of seismicity in each sequence. In this paper, we invert for focal mechanisms of small events and perform a complete fault stress state analysis of each sequence to better understand the seismic sequence evolution.

For the waveform data, we select stations within 150 km of the sequence to ensure data quality in Cushing and Guthrie. For Woodward and Fairview, the distance cutoff is increased to 250 km due to the lack of station coverage (Figure 3.1). The phase picks are provided by OGS. For events without available cataloged picks, we use an autopicker from Li and Peng (2016) to estimate P and S arrival time based on a 1D velocity model (Darold et al., 2015). The phase arrivals are refined in a pre-trained machine learning model (Ross et al., 2018). The picking quality is measured by the signal-to-noise ratio (SNR),

which is defined as the ratio between the peak absolute amplitude in the 0.5 s after and before the refined arrival time. The waveforms are then detrended and bandpassed to 1 to 20 Hz using a Butterworth filter. We use a time window of 1.5 s (0.5 s before and 1.0 s after the arrival) and 2.5 s (1 s before and 1.5 s after the arrival) for P and S wave, respectively, to compute the cross correlations in the following process.

3.3 Method

3.3.1 Earthquake Relocation and Fault Structure

We use hypoDD (Waldhauser, 2001) to compute double-difference earthquake locations using differential travel times from both phase arrivals and cross correlation. For each earthquake, we store differential times for P and S phases for up to the 100 closest neighboring earthquakes at a maximal distance of 10 km. For each earthquake pair, the travel time differences at up to 100 stations are stored. We compute the P wave cross correlations on the vertical channel and S wave cross correlation on two horizontal channels. For the S wave, the channel with higher cross-correlation coefficient is selected to measure the differential travel time. We require a minimum cross-correlation coefficient of 0.7 and use the squared cross-correlation coefficient as its weight in the relocation procedure following Schoenball and Ellsworth (2017b).

In the first set of iterations of hypoDD, we give the differential travel times from P and S phase arrivals the highest weight to constrain the overall geometry using all earthquakes. In later iterations, we add differential travel times measured from cross-correlation time lags and decrease the weight of the

catalog picks. In the last step, we use only cross-correlation measurements to resolve the fine structures. The 1D velocity model from Darold et al. (2015) is used in the relocation process. The results are shown in Figure 3.2.

In each sequence, we manually separate the events on the main fault and the secondary structures based on the earthquake relocation. For each segment structure, we use principal component analysis (Vidale & Shearer, 2006) to fit a fault plane. In this method, the eigenvalues ($\lambda_1 \geq \lambda_2 \geq \lambda_3$) and the corresponding eigenvectors (U_1, U_2, U_3) of the covariance matrix defined by earthquake epicenters are used to characterize the fault plane. The first two eigenvectors U_1, U_2 represent the surface of the fault plane, and U_3 is normal to the fault plane. The fault strike and dip angle are calculated from the direction of the normal vector U_3. A parameter planarity is defined as $1 - \lambda_3/\lambda_2$ (Vidale & Shearer, 2006), where a planarity of 1 indicates a perfectly planar shape and a planarity of 0 indicates a nearly spherical cloud of seismicity.

3.3.2 P-wave Polarity

Picking the first-motion polarities for small earthquakes is difficult due to the low signal to noise ratio. Here, we select two methods, waveform-correlation-derived relative polarities (hereinafter referred to as SVD method) (Shelly et al., 2016) and a pre-trained deep learning model for polarity classification (referred to as ML method) (Ross et al., 2018). The results from the two methods are then compared and combined to invert focal mechanisms.

SVD Method

We follow the method from Shelly et al. (2016) to derive P polarities for a whole sequence based on selected templates with known polarities. We randomly select around 100 templates in each sequence and manually pick their polarities. In this method, the waveform cross-correlation coefficients are calculated between the events and selected templates. For each event-template pair, the weighted relative polarity measurement, which is defined by the absolute difference between the peak and the secondary peak of cross correlations multiplied by the sign of the peak correlation, is used to measure the similarity between the event and template waveform.

The weighted relative polarity measurements between N events and M templates form a N by M matrix, which is then reduced to a vector by taking the left singular vector of the left unitary matrix of its Singular Value Decomposition (SVD). This vector represents the polarity pattern of all events on a specific station and channel. We measure a consistency factor by comparing the SVD-derived polarities to the manually picked polarities for the selected templates. The consistency factor w_t is defined as,

$$w_t = \sum(w_{cat} \times Pol_{svd}) / \sum(|w_{cat} \times Pol_{svd}|) \tag{3.1}$$

where w_{cat} is the manually picked polarity, and Pol_{svd} is the SVD-derived polarity. The SVD-derived polarities are then multiplied by the consistency factor to get the real polarity for a particular channel.

Machine Learning Method

In the second method, we use the pre-trained convolutional neural network (CNN) model from Ross et al. (2018) to pick the polarities. The model was trained by over two million analyst-picked polarities of earthquakes in California. We apply the same preprocessing to the waveform data as in Ross et al. (2018). The waveform is resampled to 100 Hz, detrended, and filtered between 1 and 20 Hz. Then we select a 4-s-long window centered on the P arrival time and normalize the waveform by the peak absolute amplitude in the window. The CNN model takes the 400-point time series as input and predicts the P polarities (up, down, or unknown). The results are evaluated using the metric of precision for each class of polarity (up and down). For a given class, the precision is defined as the number of true positives divided by the total number of records assigned to the class by the CNN model,

$$\begin{aligned} precison_u &= \frac{TP_{uu}}{TP_{uu} + \epsilon_{du}} \\ precison_d &= \frac{TP_{dd}}{TP_{dd} + \epsilon_{ud}} \end{aligned} \tag{3.2}$$

where u and d represents polarity up and down, TP is the number of true positives, and ϵ is the number of false positives. We test on the template events in Oklahoma, and the precision rate changes with hypocentral distance and SNR but not significantly with magnitude, consistent with the findings in California. And the precision rate for the manually picked templates is comparable to the results reported in Ross et al. (2018). The successful application of the model trained by earthquakes in California to earthquakes in Oklahoma suggests good generalization of the model.

3.3.3 Focal Mechanism Inversion

We combine the polarity results from different methods in the previous section and compute the earthquake focal mechanism solutions in hybridMT (Kwiatek et al., 2016). To ensure the quality of the inversion, we only select events with a minimum number of polarities of eight, a maximum azimuthal gap of 120°, and a maximum takeoff angle gap of 60°. Since we only use polarity information and do not include the amplitude in the input datasets, the inverted full moment tensor may not be well constrained. So we select double couple component of the moment tensor for following analysis.

We classify the focal mechanisms by rake angle following X. Chen et al. (2018). The faulting type ($f-$value) is calculated as follows,

$$f(\lambda) = \begin{cases} \lambda/90 & \text{if } |\lambda| \leq 90 \\ (180 - |\lambda|) * (\lambda/|\lambda|)/90, & \text{otherwise} \end{cases} \tag{3.3}$$

where λ is the rake angle of the nodal plane. Faulting types of -1, 0, and 1 correspond to pure normal faulting, strike-slip faulting, and reverse faulting, respectively. The rake angle continuously changes from -180° to 180°, and f-value can mitigate the gap between the rake angle of -180° and 180°.

We also use the input data resampling technique to assess the uncertainty of the moment tensor inversion. We assume 2% of input phases have wrongly picked polarity and generate 100 resampled datasets by randomly flipping the polarity of the input phase data. For each resampled input dataset, we invert a different set of focal mechanisms. We use similar criterion as in HASH program (Hardebeck & Shearer, 2008) to assign the quality of the inverted focal mechanisms. The criterion include the probability that the results from

resampled input dataset are within certain rotation angle (45°) to the preferred solution (PROB), the RMS angular difference between the preferred solution and those from resampled data (RMS_ANG), the fraction of misfit polarities (MFIT) of the preferred solution, and the station distribution (STDR) around the event. A detailed quality assignment is listed in the Table S3.1.

3.3.4 Fault Stress State

After we get the focal mechanisms in each sequence, we follow the statewide analysis in Qin et al. (2019) to calculate the fault stress state for individual events. The events are projected onto 3D Mohr circle based on the stress map in Qin et al. (2019), and a normalized parameter *understress* (Gischig, 2015) is used to quantify the criticality of the fault plane. The understress is defined as,

$$understress = (\tau_p - \tau_0)/\tau_p \tag{3.4}$$

where τ_0 is shear stress on the fault calculated from the fault geometry and stress orientations, and τ_p is shear stress at which slip initiates based on the Coulomb failure criterion under hydrostatic pore pressure. Since both τ_0 and τ_p increase linearly with depth, the defined parameter understress is independent of depth. Values of understress near 0 imply that the faults are critically stressed, while values near 1 imply negligible resolved shear stress applied on the fault, and the fault is least favorably oriented. We also calculate the parameter of excess pore pressure, which is defined as the required pore pressure increase above hydrostatic pressure for fault failure according to Mohr-Coulomb failure.

3.4 Results

In this section, we show the fault structures mapped from earthquake relocations and the focal mechanisms inverted from P first-motion polarities. Based on the focal mechanisms and regional stress field, we compute the fault stress state for individual events and present a spatiotemporal analysis of stress state for each sequence.

3.4.1 Fault Initiation and Fault Structures

Figure 3.2 shows the spatiotemporal evolution of the four sequences. The main fault in Cushing sequence is vertical and oriented in the northeast direction (N60°E). The sequence initiates around the depth of 3.8 km and migrates both upwards and downwards. The event gap (3.6 – 3.8 km) in the cross-section view is possibly due to the lack of observations at the beginning of the sequence. At the southwest end of the main fault, a short fault segment (N76°E) intersects with the main faults. A previous study has shown that Cushing sequence is in a strike-slip faulting regime with a maximum horizontal stress orientation (σ_{Hmax}) of 86° (Qin et al., 2019). Compared to the σ_{Hmax} orientation, the main fault is optimally oriented to the stress field, and the second fault segment is less optimally oriented.

The Guthrie sequence is one of the earliest reactivated sequences in Oklahoma (e.g., Benz et al., 2015; McNamara et al., 2015; Schoenball et al., 2018). The main fault has a strike of 110° and dips to the northeast. The dipping angle is nearly vertical at shallower depth (<6.5 km) and decreases to around 70° at depth. X. Chen et al. (2018) interpreted the fault structure as two subparallel faults which define an extensional jog at the changing point

of dipping angle. The fault is reactivated around the extensional jog, and the events migrate both upwards and downwards. The secondary fault is at the strike of 55° and forms a conjugate pattern to the main fault relative to the regional stress field with a σ_{Hmax} orientation of 79°. The secondary fault has ruptured most recently and is well separated from the main fault.

The Woodward sequence has been interpreted as a fault system characterized by a flower structure with strike-slip fault at deeper depth and distributed normal faults at shallower depth (Qin et al., 2018). The newly ruptured segment is at the northeast end of the main fault (N64°E), and it covers the whole depth range of the sequence from 4 km to 7 km. The secondary structures, including another parallel fault segment farther to the southwest and the intersections between the two parallel fault segments, are located at shallow depth and not well developed for the moment.

The main fault in Fairview is nearly a vertical fault with a strike of 44°. The events are migrating from northeast to southwest, away from the high injection zone to the northeast (Yeck et al., 2016). The fault is reactivated around depth of 4.5 km and migrates to deeper depth (>7 km). The secondary fault is at the strike of 97° and connected to the main fault. The fault system is located in a strike-slip regime with a σ_{Hmax} orientation of 78°. By comparing the cross-section view for Cushing and Fairview, we find that the width of fault zone ($\sim$ 0.54 km) in Fairview is larger than that in Cushing ($\sim$ 0.12 km). The width of the fault zone is defined by the distance between the 20% and 80% of seismicity distribution in across-strike direction.

3.4.2 Focal Mechanism Results

We combine the polarities results from SVD method and machine learning method. We then select events with at least eight polarity picks, a maximum of station azimuth gap of 120°, and a maximum of incident angle gap of 60° to invert for moment tensor. For Woodward and Fairview, due to the lack of station coverage, the azimuth gap cutoff is set to 100°. We obtain 305, 360, 464, and 631 focal mechanisms for Cushing, Guthrie, Woodward, and Fairview sequences, respectively. The real fault planes are differentiated from the auxiliary fault planes based on the assumption that the real fault planes are more optimally oriented relative to the regional stress field. The histograms of strike and dip angle of the selected fault planes are shown in Figure 3.3.

The strike orientations of individual fault planes are largely coherent with the inferred fault strike from seismicity. In the Cushing sequence, the dominant strike is 225 – 240°, and 69% of the fault planes are dipping steeply with a dipping angle larger than 70°. The main fault strike and the M5.0 mainshock orientation are within the strike range of the individual fault planes. The Fairview sequence has a dominant strike of 210 – 225° and a secondary strike of 300–315°, and 80% of the fault planes have a dipping angle larger than 70°. The strikes of the small events are slightly off the main fault strike mapped from seismicity. The strike difference can be explained by various orientations of small fabrics within the relatively wide fault zone. For both sequences, the dominant vertical faults planes are compatible with the vertical structure of the main faults. The slight change of strikes in the two sequences reflect the heterogeneity of the focal mechanisms in different regions of Oklahoma.

In Guthrie and Woodward, the majority of the fault planes have strike

angle in the range of [30, 60]° and [210, 240]°. Although most of the events are dipping steeply, a small subset of the events have a dip angle between 40 and 70°. The results match with the previously mapped fault structures that the dipping angle changes with depth in Guthrie and Woodward.

We classify the fault planes into different types by rake angle following X. Chen et al. (2018). The spatial distribution of the focal mechanisms in each sequence is shown in Figure 3.4. To the first order of observation, strike-slip faulting is the dominant faulting type for all sequences. In Cushing, the events on the main fault are mostly right-lateral strike-slip faulting. The intersection point of the main fault and the secondary fault structure shows the occurrence of left-lateral strike-slip faulting. In Fairview, the main fault consists of both right-lateral and left-lateral strike-slip faulting events, and the secondary fault structure hosts mainly the left-lateral strike-slip faulting events. At shallow depth, a small fraction of normal events occur in both sequences.

However, Guthrie and Woodward show a variety of faulting types on the main fault. In Guthrie, the main fault has a mixture of strike-slip and normal slip components. The normal faulting events are mainly distributed deeper than the extensional jog at a depth of 6.5 km. In Woodward, the strike-slip events are at deeper depth and normal events at shallower depth, consistent with previous results (Qin et al., 2018).

The newly ruptured segment at the depth of 6–7 km in Woodward has a dominant reverse slip component. We manually check the polarity picks and find no consistent error. A temporal distribution of the focal mechanisms is shown in Figure S3.1. The strike-slip faults are still the dominant focal mechanism type in this cluster, and the secondary type is normal faulting.

The occurrence of strike-slip faulting coincides with that of normal faulting events. However, most of the reverse events occur when the available monitoring stations are few (Figure S3.1). We also show the polarity distribution for some individual reverse faulting events in Figure S3.2. The various strike and dip angles are mostly bounded by station distributions and are not well constrained. For some events, the faulting type changes between reverse faulting and strike-slip faulting based on different resampled input datasets (Figure S3.3). With the available evidence, we argue that reverse faulting events are artifacts due to poor station coverage rather than real fault structures. For the next step, we plan to use clustering analysis on the reverse events to get more polarity observations and better constrain the inverted focal mechanisms.

3.4.3 Fault Stress State Results

With the inverted focal mechanisms, we analyze the fault stress state in a 3D Mohr circle. The regional stress field has been obtained by Qin et al. (2019). The average principal stress amplitudes in Oklahoma are 30.0, 24.84, 15.46 MPa/km, respectively. And the σ_{Hmax} orientations in Cushing, Guthrie, Woodward, and Fairview are 86°, 79°, 85°, 78°, respectively. Following the previous method (Qin et al., 2019), we project the individual fault planes in each sequence onto a 3D Mohr circle. The results are shown in Figure 3.5. We apply an understress cutoff of 0.15 to identify optimally oriented faults. For sequences with a relatively uniform strike-slip faulting mechanism in Cushing and Fairview, the percentage of optimally oriented faults is the high (75% and 72%). In Fairview, although the M5.1 main shock is not optimally oriented to the local stress field (Qin et al., 2019), many smaller events are more

critically stressed. This can be explained by the observation that small fabrics of different orientations exist in the fault zone, and the optimally oriented fabrics are ruptured. Guthrie has 59% optimally oriented events. Woodward has the lowest percentage partially due to the normal and reverse faulting events.

The spatial distribution of the fault stress state is shown in Figure 3.6. In Cushing, most of the events on the main fault are optimally oriented. The non-optimally oriented faults only occur on normal faulting events at shallow depth of the main fault and the intersection between the main fault and the secondary fault. In Fairview, the small events on the main fault are mostly optimally oriented with the exception of some shallow events. Different from Cushing, many events on the secondary fault structures are also optimally oriented, which could imply potential earthquake hazard on the secondary fault segment. In Guthrie, fewer events are critically stressed than Cushing and Fairview due to the normal faulting events. The main fault in Woodward is the least critically stressed of the four sequences. The strike-slip events on the main fault are mostly optimally oriented, and the normal events at shallow depth are less optimally oriented. The rupture of the less optimally oriented normal events have been attributed to the earthquake interactions (Qin et al., 2018). The newly ruptured events at depth of 6–7 km are the least optimally oriented. The reverse events are possibly artifacts because of the lack of station coverage (Figure S3.1 to S3.3).

Figure 3.7 show the temporal evolution of the four sequences. Cushing is featured by two foreshock-mainshock-aftershock subsequences. The dominant faulting type is strike-slip, and some normal events occur after the M5.0 main

shock. Each subsequence starts with larger events that are optimally oriented (understress<0.15). In the following stage, some less optimally oriented events rupture, which are possibly triggered by earlier events. Guthrie is characterized as a swarm sequence without significantly large events. The sequence also starts with optimally oriented events, and the majority of the events have understress values between 0 and 0.5. Few non-optimally oriented events are triggered.

The Woodward sequence is incomplete due to the lack of observations at the beginning of the sequence. The events with large understress at the beginning and end of the sequence are possibly not well constrained due to station coverage. We get the stress state for part of the sequence from mid 2016 to 2018, when temporary stations and industrial stations are available. Similar to Guthrie, the Woodward sequence does not show large variations in stress state and event magnitude.

The Fairview sequence initiated later than Woodward, and the focal mechanism solutions cover the majority of the sequence except for the beginning. The M5.1 mainshock is less optimally oriented, followed by both optimally oriented events and non-optimally oriented events. The occurrence of non-optimally oriented events possibly suggests the existence of earthquake interactions.

3.5 Discussion

In this section, first we include more details about the applications of the two different polarity-picking methods. We discuss the validity and applicability of each method by comparison. Then we examine the uncertainties of the inverted focal mechanisms and their influence on fault classification and stress

state calculation. Finally, we combine all the findings, including fault structure, stress state, and pore pressure to evaluate the potential earthquake hazards.

3.5.1 Validity of the Polarity-picking Methods

In the SVD polarity method, the relative polarities are obtained by comparing all events to the selected templates using cross correlation. To control the quality of the results, we apply a cross-correlation coefficient cutoff of 0.8, and search for a cutoff value for SVD results by requiring the derived polarities for templates are over 98% consistent to the manually picked polarities. The results are shown in Table S3.3.

For ML polarity methods, we iteratively search for a hypocentral distance cutoff and a SNR cutoff to ensure that the precision of templates predictions is over 98%. The results are also shown in Table S3.3. Without any adjustment of the machine learning model, the prediction accuracy can reach comparable level to the test results in the paper (Ross et al., 2018). The results suggest that the model has good generality to be applied to a different region. Considering we are only testing on a small dataset with relatively high SNR, we might need to retrain the model with a large dataset of Oklahoma earthquakes in the future for general purpose.

By comparison, over 96% of the polarities for common events from SVD and ML methods are consistent, which validates the applications of the two polarity-picking methods. For the SVD method, we find that the polarities of some large events (M$>$ 4) are not well resolved, because the correlation coefficients between relatively large events and small events are low. So the SVD method is most applicable to similar event clusters. On the contrary, the

ML model can be applied to events with more complex sources. Both methods are affected by low SNR. For SVD methods, only events with similar waveform to the templates can be resolved, and the noisy waveform will decrease the cross-correlation coefficients. For ML methods, the model can classify the noisy waveform as polarity unknown if SNR is low.

3.5.2 Focal Mechanism Uncertainties

During the focal mechanism inversion, we resampled the input data 100 times and inverted for the focal mechanisms for each resampling. The results are used to examine the uncertainties of the focal mechanisms. For each event, we compute the rotation angles between the FM from the complete input dataset and the FMS from the resamplings. A distribution of the mean rotation angle is shown in Figure S3.4. The mean rotation angles for Cushing, Guthrie, Woodward, and Fairview are 18°, 22°, 26°, and 22°, respectively. The rotation angle for Cushing sequence is the smallest, suggesting the focal mechanisms are best constrained. The opposite is true for Woodward sequence. As shown in Table S3.2, most of the focal mechanisms are of A quality, and fewer than one third of the focal mechanisms in Woodward are of A quality.

The fault rake angle determines the faulting type, and the f-value, as a function of rake angle, is used to represent the faulting type. For each event, we extract all rake angles from the resampling results and calculate the standard deviations of their f-values. Figure S3.5 shows the distribution of standard deviations of f-value for each sequence. The four sequences have similar deviation value around 0.27. Based on the definition in Equation 3.3, the f-value of 0.27 is equivalent to a rake angle change of 24°.

The fault strike and dip angles combined control the stress state. To qualify the influence of the uncertainties, we check the standard deviation values of understress value of the inversion results from resamplings. The results are shown in Figure S3.6. The mean standard deviation of understress value is 0.12, 0.14, 0.16, and 0.15 for Cushing, Guthrie, Woodward, and Fairview, respectively. Even though the FMS are not very well constrained in Woodward compared to the other three sequences, it does not severely influence the stress state.

3.5.3 Potential Earthquake Hazard

The fault orientations in pressurized regions influence the rupture propagation behaviors of the faults. The rupture behavior then influences the possible maximum magnitude of earthquakes, which is essential to evaluate the potential earthquake hazards. Gischig (2015) studied the fault rupture behaviors using rupture simulations and found that different orientation of the pressurized faults might results in uncontrolled rupture-front propagation beyond the pressure or rupture-front propagation arresting at the pressure front. Thus, the largest possible magnitude was determined either by fault properties or injection volumes. Figure 3.8 shows the cross plot of understress and modeled pore pressure for the four sequences. The understress values represent the criticality of the fault in the regional stress field. Instead of the injection pressure used in (Gischig, 2015), for y-axis we use the calculated pore pressure from hydrogeologic modeling (Zhai et al., 2019).

In Figure 3.8, Cushing and Fairview have more optimally oriented faults, as shown in Figure 3.5. Cushing sequence is in the highest pore pressure region

with a mean value of 0.89 MPa. Combined with the previous observation that the main fault in Cushing is a relatively simple vertical planar fault, it is possible that uncontrolled rupture occurs and results in a maximum magnitude of 5.0, which is comparable to the magnitude size determined by fault size (Qin et al., 2019). Besides the largest magnitude, more moderate earthquakes ($5 > M \geq 3$) occur on optimally oriented faults in the Cushing sequence. In Fairview, the pore pressure is around 0.15–0.3 MPa. The abundance of optimally oriented faults and relatively high pore pressure put the sequence at high risk. Compared to the empirical relationship between fault length and magnitude for natural earthquakes (Wells & Coppersmith, 1994), the maximum magnitude M5.1 is slightly lower than the prediction (Qin et al., 2019). This observation suggests that the rupture of the mainshock in Fairview probably is still controlled by pressure-front due to its non-optimal orientation.

In Guthrie, the sequence hosts a smaller percentage of optimally oriented events compared to Cushing. The modeled pore pressure (mean pore pressure 0.26 MPa) is much lower than Cushing. Considering the relatively complex fault structures, the fault is most probably controlled by the pressurized front from injection and does not host a significantly large earthquake. In Woodward, due to the smallest percentage of optimally oriented events and the lowest pore pressure increase (mean value of 0.12 MPa), the event magnitude is relatively small compared to other sequences. The above observations show that various factors, including fault structures, fault orientations, and injections, influence the fault rupture and seismic hazard evaluation. Overall, the potential seismic hazard is higher for Cushing and Fairview sequences.

3.6 Conclusions

We map the fault structures, invert focal mechanisms, and compute the individual fault stress states for four selected clusters in Oklahoma. The main faults in Cushing and Fairview are vertical planar faults. The dominant focal mechanism type is strike-slip faulting and over 70% of the individual events are optimally oriented. On the other hand, Guthrie and Woodward show more complex fault structures with a mixture of strike-slip and normal faulting components. In the regional stress field, Guthrie and Woodward have a lower percentage of optimally-oriented faults, 59% and 47%, respectively. As a result, Cushing and Fairview are more susceptible to uncontrolled rupture process and have a higher potential seismic hazard level. Guthrie and Woodward have lower seismic hazard level, and the largest earthquake magnitude is smaller than 4.5. Our results show heterogeneous fault structures and focal mechanisms in induced sequences. Multiple factors, including the fault structures, fault stress state, and injection histories, influence the rupture process and seismic hazard evaluation.

References

Benz, H. M., McMahon, N. D., Aster, R. C., McNamara, D. E., & Harris, D. B. (2015). Hundreds of earthquakes per day: The 2014 guthrie, oklahoma, earthquake sequence. *Seismological Research Letters*, *86*(5), 1318–1325.

Chen, X., Haffener, J., Goebel, T. H., Meng, X., Peng, Z., & Chang, J. C. (2018). Temporal correlation between seismic moment and injection volume for an induced earthquake sequence in central oklahoma. *Journal of Geophysical Research: Solid Earth*, *123*(4), 3047–3064.

Chen, X., Nakata, N., Pennington, C., Haffener, J., Chang, J. C., He, X., ... Walter, J. I. (2017). The pawnee earthquake as a result of the interplay among injection, faults and foreshocks. *Scientific reports*, *7*(1), 1–18.

Chen, Y., & Huang, L. (2019). Adaptive moment-tensor joint inversion of clustered microseismic events for monitoring geological carbon storage. *Geophysical Journal International*, *219*(1), 80–93.

Darold, A. P., Holland, A. A., Morris, J. K., & Gibson, A. R. (2015). Oklahoma earthquake summary report 2014. *Okla. Geol. Surv. Open-File Rept. OF1-2015*, 1–46.

De Matteis, R., Convertito, V., & Zollo, A. (2016). Bistrop: Bayesian inversion of spectral-level ratios and p-wave polarities for focal mechanism determination. *Seismological Research Letters*, *87*(4), 944–954.

Ellsworth, W. L. (2013). Injection-induced earthquakes. *Science*, *341*(6142), 1225942. doi: 10.1126/science.1225942

Gischig, V. S. (2015). Rupture propagation behavior and the largest possible earthquake induced by fluid injection into deep reservoirs. *Geophysical Research Letters*, *42*(18), 7420–7428.

Goebel, T., Weingarten, M., Chen, X., Haffener, J., & Brodsky, E. (2017). The 2016 mw5. 1 fairview, oklahoma earthquakes: Evidence for long-range poroelastic triggering at¿ 40 km from fluid disposal wells. *Earth and Planetary Science Letters*, *472*, 50–61.

Hardebeck, J. L., & Shearer, P. M. (2008). *Hash: A fortran program for computing earthquake first-motion focal mechanisms–v1. 2–january 31, 2008.*

Keranen, K. M., Savage, H. M., Abers, G. A., & Cochran, E. S. (2013). Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 Mw 5.7 earthquake sequence. *Geology*, *41*(6), 699–702. doi: 10.1130/G34045.1

Keranen, K. M., Weingarten, M., Abers, G. A., Bekins, B. A., & 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

Kwiatek, G., Martínez-Garzón, P., & Bohnhoff, M. (2016). Hybridmt: A matlab/shell environment package for seismic moment tensor inversion and refinement. *Seismological Research Letters*, *87*(4), 964–976.

Li, Z., & Peng, Z. (2016). An automatic phase picker for local earthquakes with predetermined locations: Combining a signal-to-noise ratio detector with 1d velocity model inversion. *Seismological Research Letters*, *87*(6),

1397–1405.

McGarr, A., & Barbour, A. J. (2017). Wastewater disposal and the earthquake sequences during 2016 near fairview, pawnee, and cushing, oklahoma. *Geophysical Research Letters*, *44*(18), 9330–9336.

McNamara, D. E., Benz, H. M., Herrmann, R. B., Bergman, E. A., Earle, P., Holland, A., ... Gassner, A. (2015). Earthquake hypocenters and focal mechanisms in central oklahoma reveal a complex system of reactivated subsurface strike-slip faulting. *Geophysical Research Letters*, *42*(8), 2742–2749.

Pugh, D., White, R., & Christie, P. (2016). A bayesian method for microseismic source inversion. *Geophysical Journal International*, *206*(2), 1009–1038.

Qin, Y., Chen, X., Carpenter, B. M., & Kolawole, F. (2018). Coulomb stress transfer influences fault reactivation in areas of wastewater injection. *Geophysical Research Letters*, *45*(20), 11–059.

Qin, Y., Chen, X., Walter, J. I., Haffener, J., Trugman, D. T., Carpenter, B. M., ... Kolawole, F. (2019). Deciphering the stress state of seismogenic faults in oklahoma and southern kansas based on an improved stress map. *Journal of Geophysical Research: Solid Earth*, *124*(12), 12920–12934.

Ross, Z. E., Meier, M.-A., & Hauksson, E. (2018). P wave arrival picking and first-motion polarity determination with deep learning. *Journal of Geophysical Research: Solid Earth*, *123*(6), 5120–5129.

Schoenball, M., & Ellsworth, W. L. (2017a). A systematic assessment of the spatiotemporal evolution of fault activation through induced seismicity in oklahoma and southern kansas. *Journal of Geophysical Research: Solid Earth*, *122*(12), 10–189.

Schoenball, M., & Ellsworth, W. L. (2017b). Waveform-relocated earthquake

catalog for oklahoma and southern kansas illuminates the regional fault network. *Seismological Research Letters*, *88*(5), 1252–1258.

Schoenball, M., Walsh, F. R., Weingarten, M., & Ellsworth, W. L. (2018). How faults wake up: the guthrie-langston, oklahoma earthquakes. *The Leading Edge*, *37*(2), 100–106.

Shelly, D. R., Hardebeck, J. L., Ellsworth, W. L., & Hill, D. P. (2016). A new strategy for earthquake focal mechanisms using waveform-correlation-derived relative polarities and cluster analysis: Application to the 2014 long valley caldera earthquake swarm. *Journal of Geophysical Research: Solid Earth*, *121*(12), 8622–8641.

Skoumal, R. J., Kaven, J. O., & Walter, J. I. (2019). Characterizing seismogenic fault structures in oklahoma using a relocated template-matched catalog. *Seismological Research Letters*. doi: 10.1785/0220190045

Vavryčuk, V. (2015). Inversion for the composite moment tensor. *Bulletin of the Seismological Society of America*, *105*(6), 3024–3035.

Vidale, J. E., & Shearer, P. M. (2006). A survey of 71 earthquake bursts across southern California: Exploring the role of pore fluid pressure fluctuations and aseismic slip as drivers. *Journal of Geophysical Research: Solid Earth*, *111*(5), 1–12. doi: 10.1029/2005JB004034

Waldhauser, F. (2001). hypodd—a program to compute double-difference hypocenter locations (hypodd version 1.0-03/2001). *US Geol. Surv. Open File Rep., 01*, *113*.

Walter, J. I., Ogwari, P., Thiel, A., Ferrer, F., Woelfel, I., Chang, J. C., ... Holland, A. A. (2020). The oklahoma geological survey statewide seismic network. *Seismological Research Letters*, *91*(2A), 611–621.

Wells, D. L., & Coppersmith, K. J. (1994). New empirical relationships among

magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the seismological Society of America*, *84*(4), 974–1002.

Wu, Q., Chen, X., & Abercrombie, R. E. (2019). Source complexity of the 2015 mw 4.0 guthrie, oklahoma earthquake. *Geophysical Research Letters*, *46*(9), 4674–4684.

Yeck, W. L., Weingarten, M., Benz, H. M., McNamara, D. E., Bergman, E., Herrmann, R., ... Earle, P. S. (2016). Far-field pressurization likely caused one of the largest injection induced earthquakes by reactivating a large preexisting basement fault structure. *Geophysical Research Letters*, *43*(19), 10–198.

Zhai, G., Shirzaei, M., Manga, M., & Chen, X. (2019). Pore-pressure diffusion, enhanced by poroelastic stresses, controls induced seismicity in oklahoma. *Proceedings of the National Academy of Sciences*, *116*(33), 16228–16233.

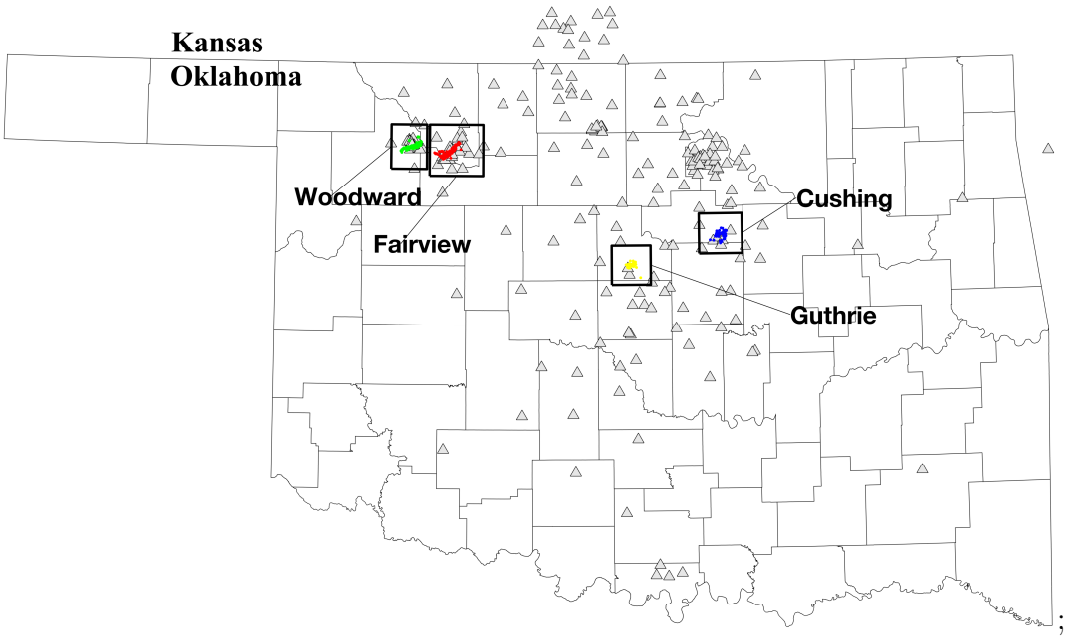

Figure 3.1: Stations (gray triangles) and earthquakes (colored circles in four sequences, 2009–2018) used in this study. The inset figure shows the location of the study region.

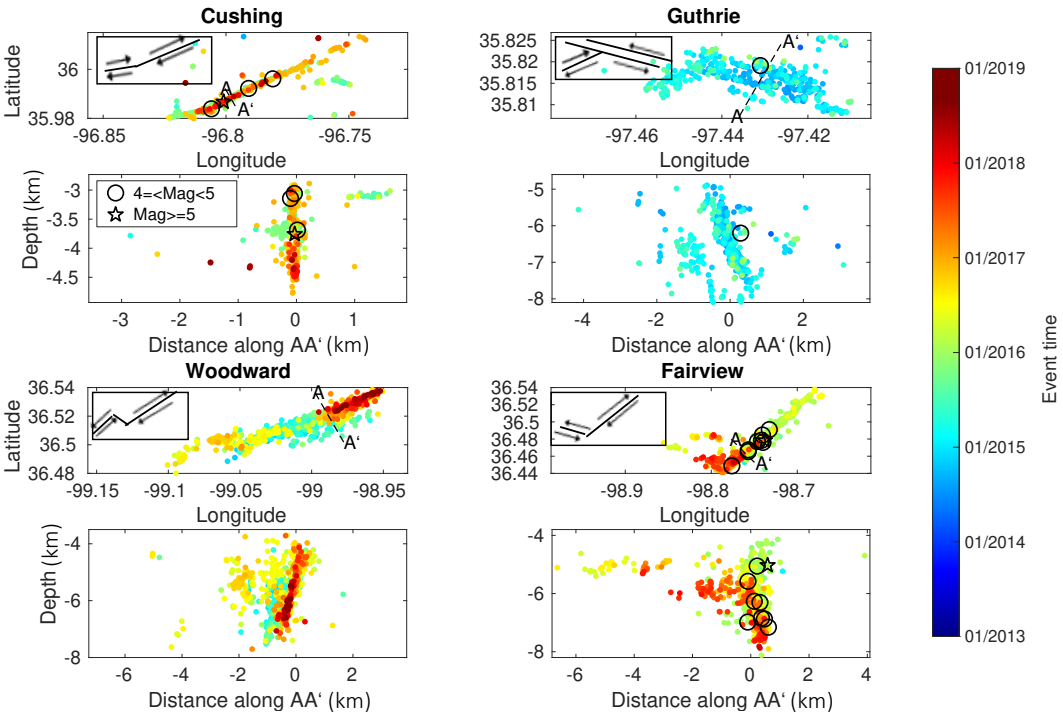

Figure 3.2: Earthquake relocations for four sequences. The top panels show the map view, and the bottom panels show the cross-section view (along AA′). The events are colored by event origin times. Star: M≥5 event; circles: M≥4 events. The inset diagrams on the map view show the interpretation of the fault structure. The diagrams are not to scale.

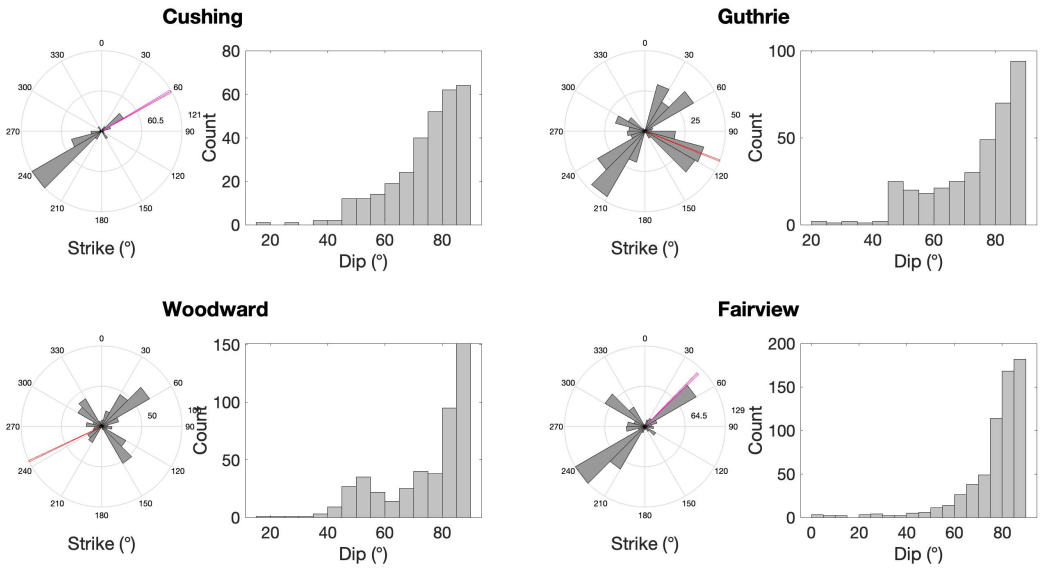

Figure 3.3: Histogram of strike and dip angles of focal mechanisms for the four sequences. The red lines show the orientation of the main faults interpreted from seismicity. For Cushing and Fairview, the magenta lines show the strike of M≥5 earthquakes.

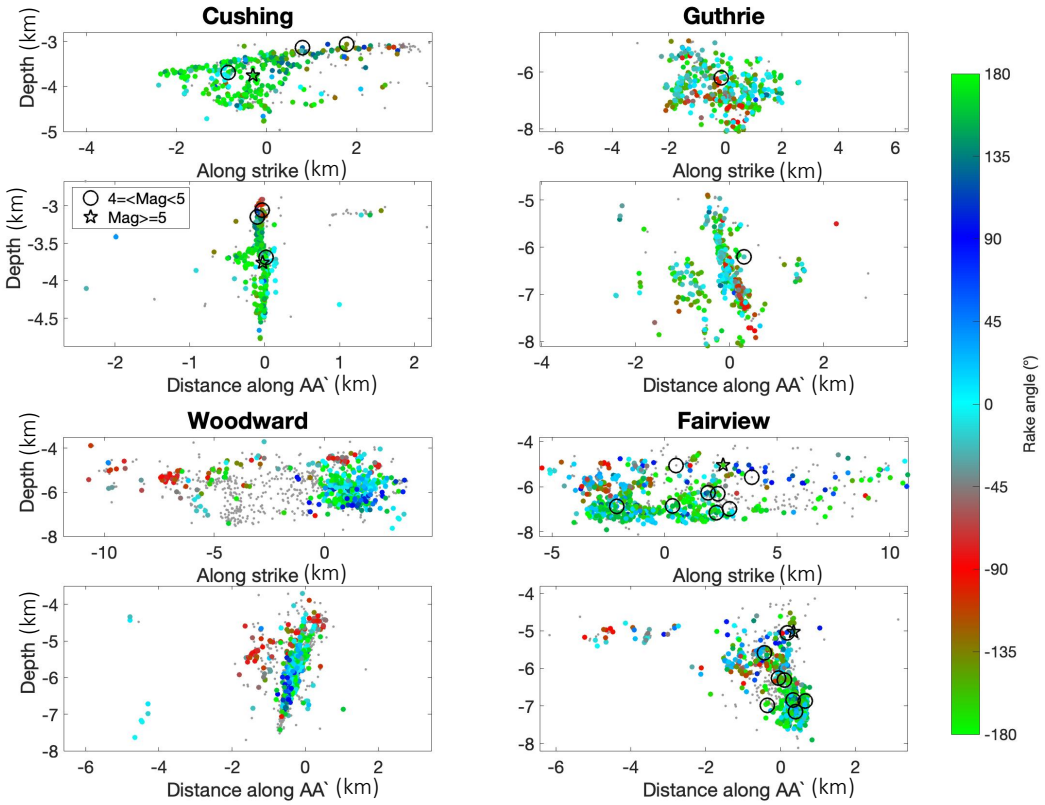

Figure 3.4: Focal mechanism solutions for four sequences. The top panels show the cross-section view along main fault strike, and the bottom panels show the cross-section view (along AA′). The events are colored by fault classes. S: strike-slip, N: normal, R: reverse. Star: M≥5 event; circles: M≥4.

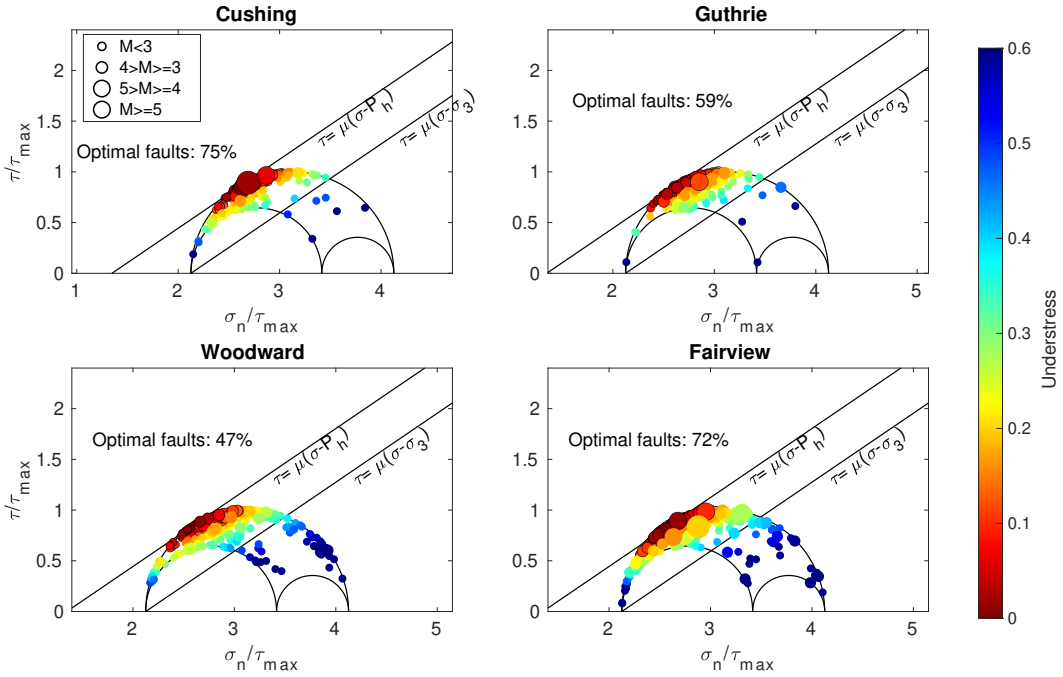

Figure 3.5: The stress state of individual fault planes in each sequence in 3D Mohr diagram. The three semi-circles represent the stress tensor, and the two straight lines represent the fault strength under hydrostatic fluid pressure and lithostatic pressure assuming a friction coefficient of 0.68.

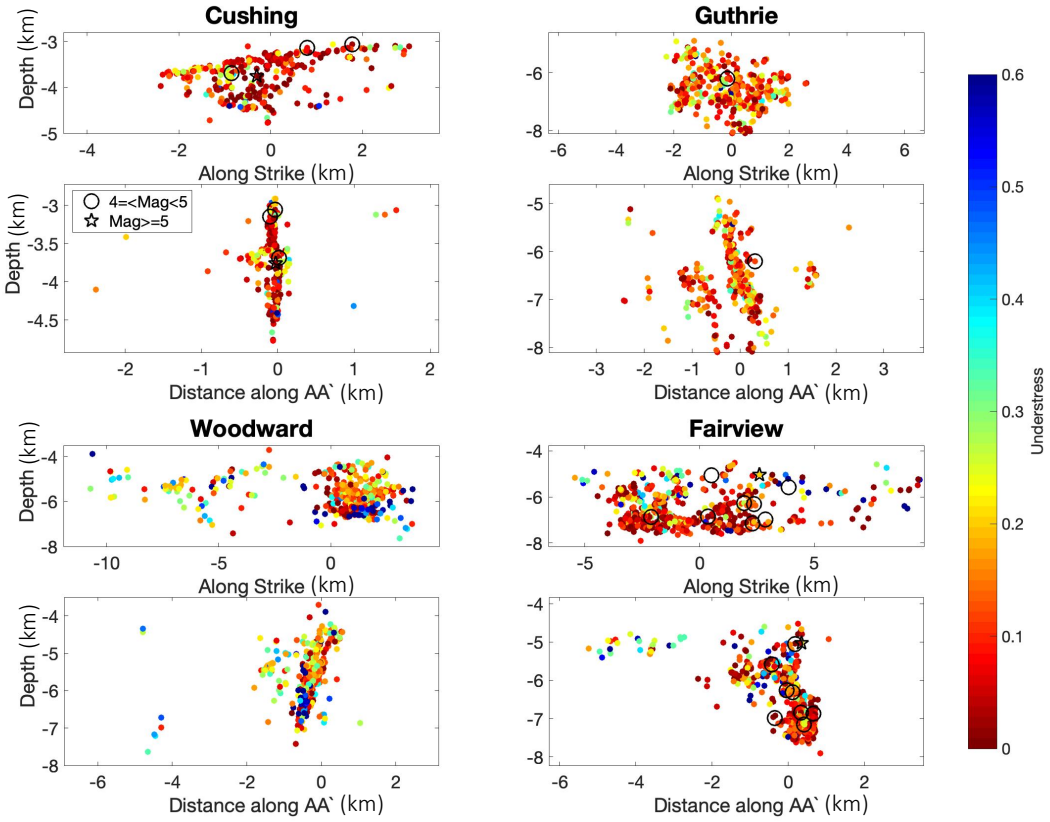

Figure 3.6: Fault stress state for four sequences. The top panels show the cross-section view along strike, and the bottom panels show the cross-section view (along AA′). The events are colored by understress values. Star: M≥5 event; circles: M≥4.

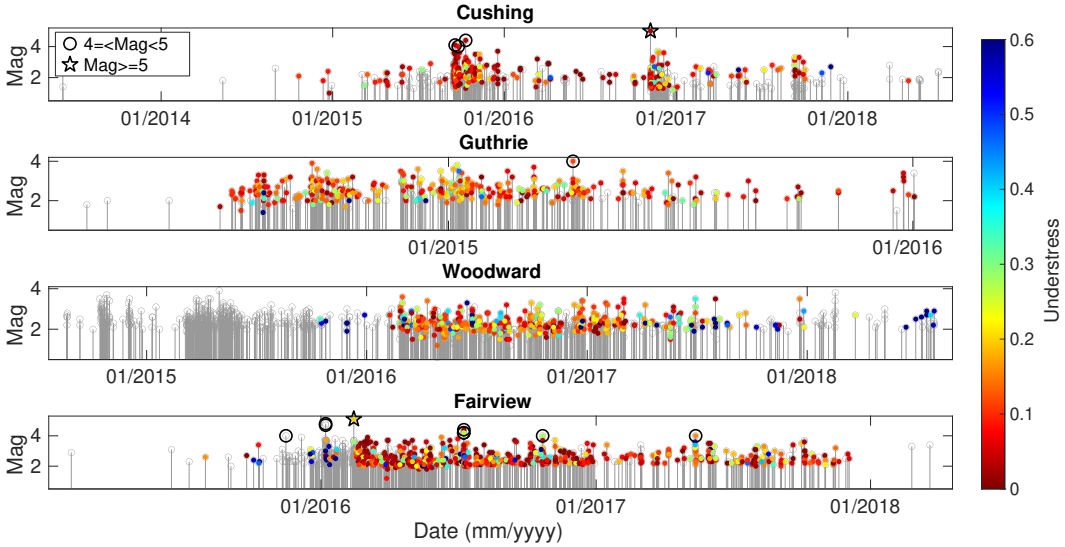

Figure 3.7: Temporal evolution of the four sequences. The top panel shows the event magnitude with time (black and gray for events with and without inverted focal mechanisms). The middle panel shows the focal mechanism classes with time. The bottom panel shows the stress state of individual fault planes with time. The gray filled events are focal mechanisms of quality D. Star: M≥5 event; circles: M≥4.

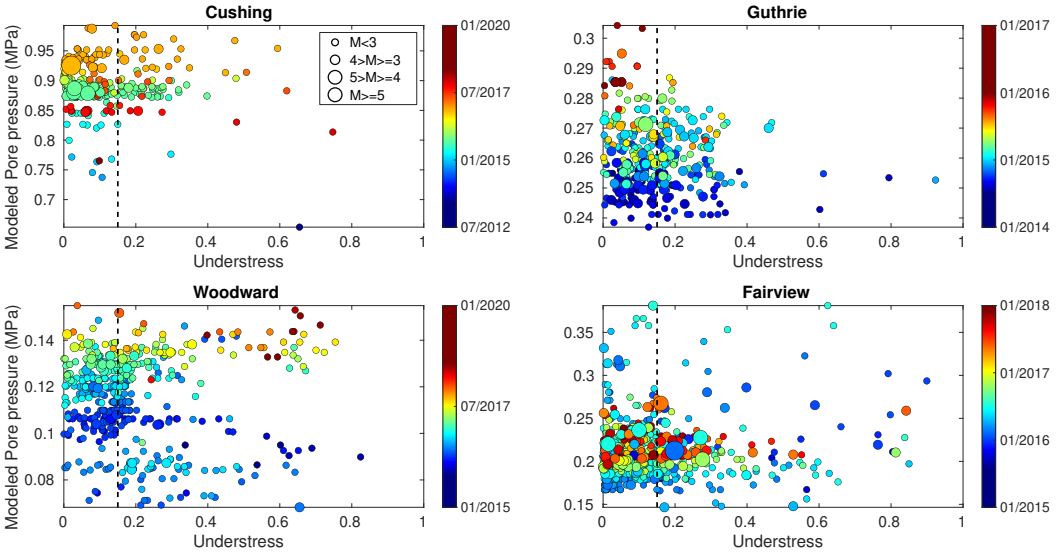

Figure 3.8: Cross plot of understress and modeled pore pressure for each sequence. The events are colored by event time and scaled by magnitude. The dashed lines denote the understress value of 0.15. The pore pressure data are from Zhai et al. (2019).

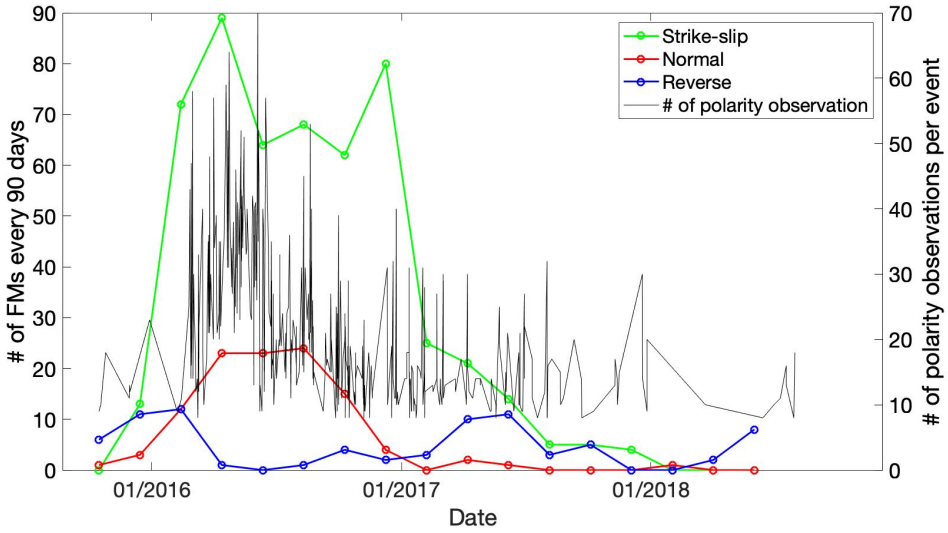

Figure S3.1: The temporal distribution of focal mechanisms and the number of available stations for Woodward sequence.The reverse events mainly occur at the beginning and end of the sequence.

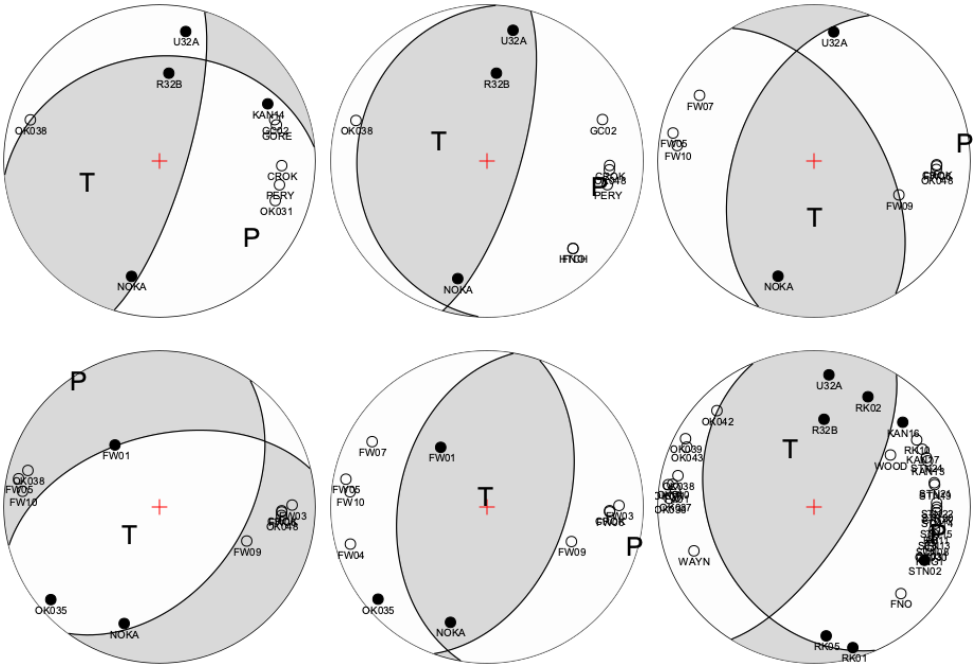

Figure S3.2: Focal mechanism solutions with observed polarities for some selected reverse events in Woodward sequence.

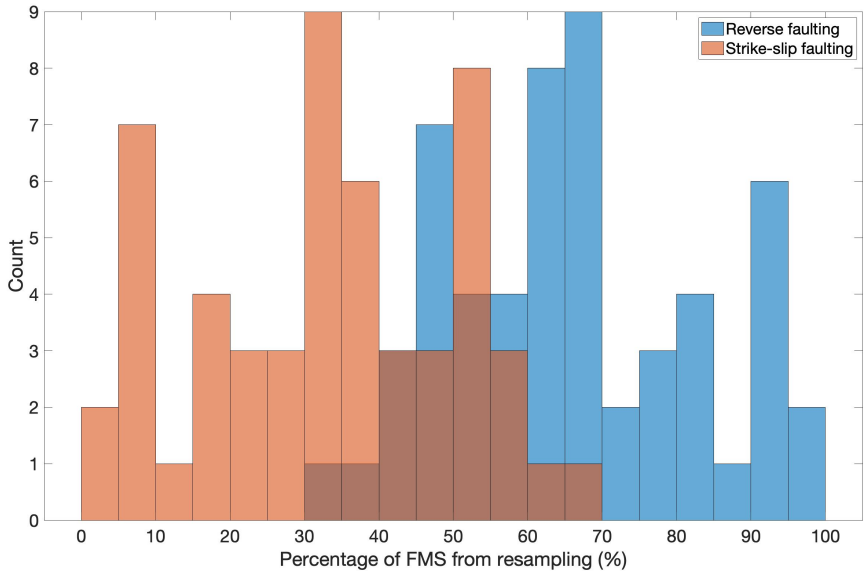

Figure S3.3: The faulting type inverted from resampled input dataset for the reverse events in Woodward.

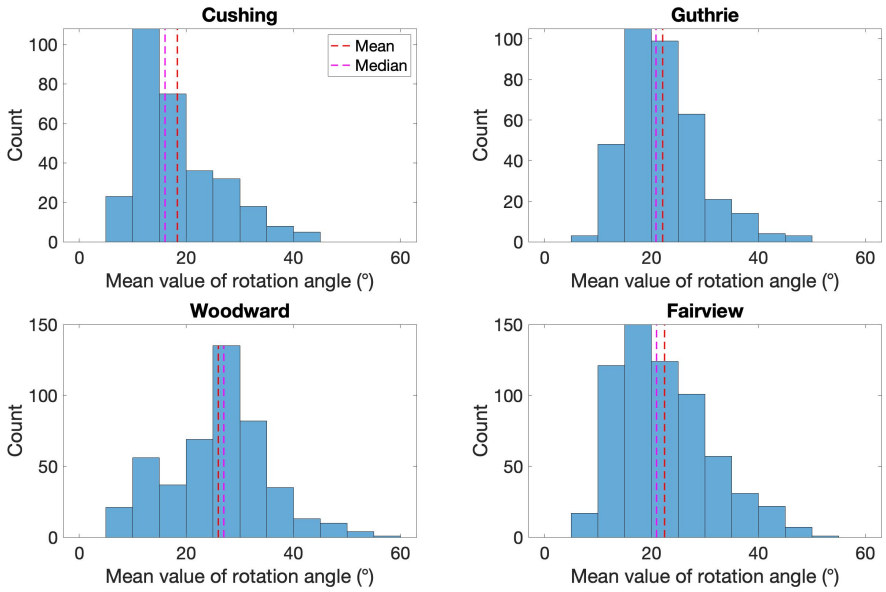

Figure S3.4: Distribution of the mean rotation angle between the FMS from complete input dataset and those from 100 resamplings.

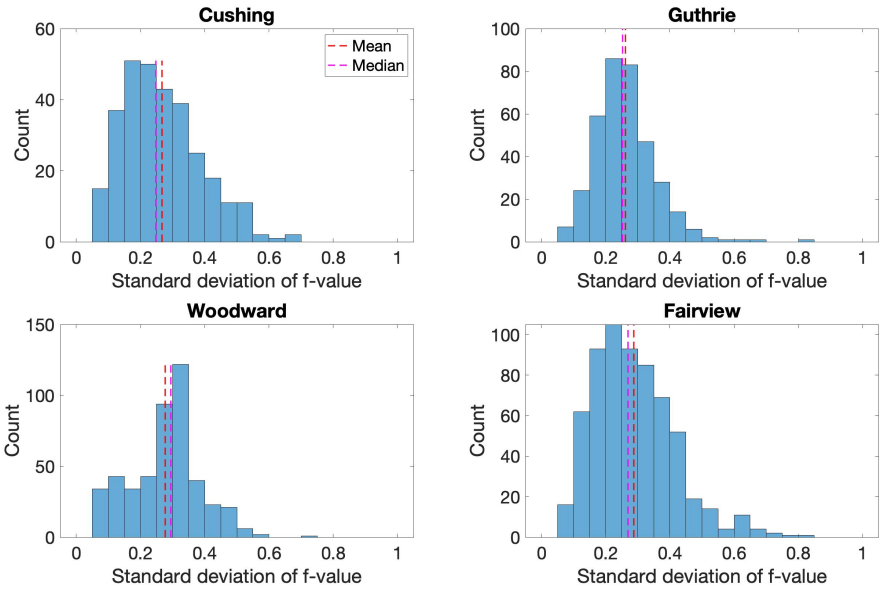

Figure S3.5: Distribution of the standard deviations of f-value calculated using FMS inverted from 100 resamplings.

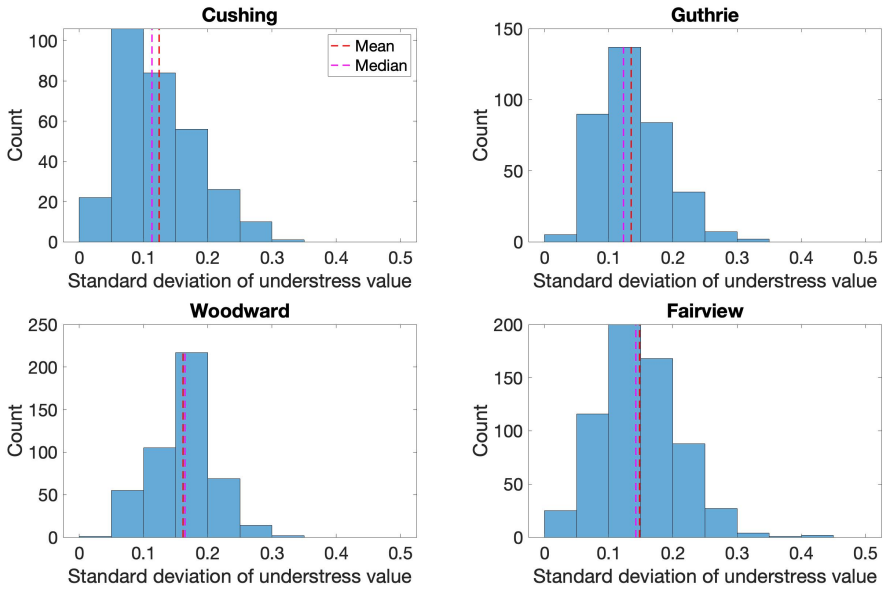

Figure S3.6: Distribution of the standard deviations of understress value calculated using FMS inverted from 100 resamplings.

Table S3.1: Quality characterization of inverted focal mechanisms.

PROB	RMS_ANG	MFIT	STDR	Quality
≥ 0.8	≤ 25	≤ 0.15	≥ 0.5	A
≥ 0.6	≤ 35	≤ 0.2	≥ 0.4	B
≥ 0.5	≤ 45	≤ 0.3	≥ 0.3	C
< 0.5	> 45	> 0.3	< 0.3	D

Table S3.2: Quality characterization results of inverted focal mechanisms.

Sequence	A	B	C	D	Total
Cushing	228	47	15	15	305
Guthrie	289	62	4	5	360
Woodward	149	249	46	19	463
Fairview	352	156	80	33	631

Table S3.3: Polarity result comparison from SVD and ML methods.

Sequence	Temp	SVD Pols	ML Pols	Consist	Inconsist	Consist rate
Cushing	2807	9031	12212	7025	178	97.53%
Guthrie	2466	9954	14008	8336	168	98.02%
Woodward	4024	8638	13682	4959	67	98.67%
Fairview	4915	12301	26065	9864	321	96.85%

Chapter 4

Coulomb Stress Transfer Influences Fault Reactivation in Areas of Wastewater Injection

Plain Language Summary

The earthquakes in wastewater injection areas have been mainly linked to fluid injection, which increases the pore pressure or poroelastic stress and promotes fault failure. Only limited studies have explored another possible driving mechanism – stress interactions between the earthquakes during the fault reactivation in those areas. In this study, we focus on an isolated earthquake cluster in the northwest Oklahoma, a wastewater injection area and study how the earthquake interactions influence the step-by-step reactivation of the fault system. The calculated stress interactions of small earthquakes on the fault planes are larger than the pore pressure change and at least comparable to the

poroelastic stress change from fluid injection. Our results suggest that the fluid injection is not the only driving mechanism of seismicity in wastewater injection areas, and earthquake interactions should also be considered for mitigating induced seismicity.

4.1 Introduction

Much of the seismicity in Oklahoma is highly clustered in space, which can be used to map faults in the subsurface and study triggering mechanisms of earthquake sequences (e.g., X. Chen et al., 2017; Keranen et al., 2013; Schoenball & Ellsworth, 2017a; Yeck et al., 2017). In northwestern Oklahoma, a cluster, isolated from the main seismicity area, initiated near the city of Woodward in 2014 (hereinafter referred to as "Woodward cluster"). Stress inversion results show that the Woodward area is located within a transtensional stress field, rather than the dominant strike-slip faulting regime of central Oklahoma (Qin, 2017). Most of the earthquakes in the cluster are in the crystalline basement with depth deeper than 4km. The temporal evolution of the cluster resembles a typical swarm-type sequence without any events larger than magnitude 4.0. Relocated catalogs from Schoenball and Ellsworth (2017b) and C. Chen (2016) include the initiation portion of the cluster and provide sufficient data (Figure 4.1a-f) to map fault structures and the reactivation process.

In the brittle crust, strike-slip systems tend to develop complex arrays of structures with minor faults striking obliquely to the overall trend (e.g., Le Guerroué & Cobbold, 2006). In cross-section view, the fault tends to be steep at depth and splays upward, forming characteristic flower structures (Sylvester,

1988). Those structures can be readily identified in two-dimensional (2D) or three-dimensional (3D) seismic profiles (e.g., Harding, 1985). The structures have been observed in seismic reflection profiles in Oklahoma (e.g., Figure S4.1). Due to the unavailability of seismic reflection data in the Woodward area, we rely on accurately relocated earthquakes to map fault structures (Figure 4.1).

To identify the mechanism of the fault reactivation process, we focus on the spatiotemporal distribution of events in the sequence. A purely fluid induced earthquake sequence can be simulated by the Poisson process, where successively occurring events are not causally related to each other (Langenbruch et al., 2011). The nearest neighbor approach has been used to differentiate induced and natural seismicity at the Coso Geothermal Field (Schoenball et al., 2015) and other regions (Zaliapin & Ben-Zion, 2015). We use this approach to study the characteristics of earthquakes and the resulting implications for the mechanism of fault reactivation.

In Oklahoma, wastewater disposal has been the main driving mechanism for recent seismicity since 2009 (e.g., Ellsworth, 2013; Keranen et al., 2013; McGarr, 2014). Concerning the triggering mechanisms, pore fluid pressure change is more prominent at small distances, whereas at large distances, poroelastic stress changes may surpass the pore pressure change (e.g., Goebel et al., 2017; Segall & Lu, 2015). Goebel et al. (2017) calculated the pore pressure change and poroelastic stress change in Woodward area including both the nearby well and wells in the high-rate injection zone $\sim$ 40km to the east of the cluster. The median pore pressure and poroelastic pressure changes are 0.03MPa and 0.12MPa, respectively. We compare the Coulomb stress from earthquake interactions to fluid-related pressure changes to better understand

the driving forces of fault reactivation.

4.2 Data

4.2.1 Earthquake Relocations

To study fault structures, we use the relocated catalog from Schoenball and Ellsworth (2017b). They used waveform data from a combination of public and industry-operated seismic networks and applied the double-difference relative location method ("hypoDD") (Waldhauser & Ellsworth, 2000) with differential times measured from waveform cross correlation. The catalog includes 418 events from August 2014 to November 2016 in Woodward with magnitude from 1.5 to 3.8. The relative location precision is on the order of 10m in the horizontal direction and about 50m in the vertical direction with a magnitude of completeness of 2.5. Figure 4.1a and 4.1b show the relocations of the Woodward cluster in map view and cross-section view.

Since the relocated catalog from Schoenball and Ellsworth (2017b) is based on waveform cross correlation and the station distribution was sparse in that area before 2015, some events could not be relocated by hypoDD. To analyze the fault reactivation process with spatiotemporal distribution of earthquakes, we need a more complete catalog. We refer to the catalog by C. Chen (2016), which was obtained using the double difference method and a 3D velocity model derived from Oklahoma earthquakes based on differential times from phase picks in the Oklahoma Geological Survey (OGS) catalog. The location uncertainty is $\sim$ 40m horizontally and $\sim$ 100m vertically. The relocation results are slightly dispersed in comparison to Schoenball and Ellsworth (2017b). However, the

catalog includes 763 events from August 2014 to October 2016 with magnitude from 1.1 to 3.8 and a magnitude of completeness of 2.5. The number of the relocated events is comparable to that in the raw catalog from OGS. This catalog is used in the analysis of the fault reactivation process (Figure 4.2 and Figure 4.3).

4.2.2 Focal Mechanism Solutions

For focal mechanism solutions in Woodward, we use a subset of 102 events in the focal mechanism solution catalog from the Oklahoma Geological Survey (OGS) from December 2014 to November 2016. The date range covers similar period as the relocated catalogs from Schoenball and Ellsworth (2017b) and C. Chen (2016). The focal mechanism solution catalog was obtained based on first motion polarities using either FPFIT (Reasenberg & Oppenheimer, 1985) or HASH (Hardebeck et al., 2005) programs, which are built-in functions from the SEISAN package (Windows, 2009) used for routine processing. We observe a mix of strike-slip faulting and normal faulting in the cluster (Figure 4.1c and Figure S4.2). The focal mechanism distribution in AA$'$ cross-section view (Figure 4.1d) shows that at deeper depth ($>$ 7.7km), the faulting type is dominated by strike-slip faults; while above 7.7km, the faulting type transitions to oblique faulting with mixed normal and strike-slip components. Combining the information from Figure 4.1b and 4.1d, we find that the structures at depth are characterized by a strike-slip fault, and the structures above are characterized by oblique normal faults.

4.3 Fault Interpretation

Using the seismicity distribution in Figure 4.1, we map the fault system of the Woodward cluster. In the map view (Figure 4.1e), we separate the seismicity into three groups with different distribution trends. Then, for each group, the fault strike is calculated from the seismicity distribution using principal component analysis (Vidale & Shearer, 2006) and the fault dip is the average value of the dip angles of the focal mechanism solutions in that group. We note that the dip angle has strong influence on the resulting Coulomb stress amplitude, which will be further discussed. Table 4.1a lists the fault geometries and locations. The depth ranges of the three faults are obtained from the 15th and 85th percentile of the depth distribution from all the earthquakes within each fault. R1 is the main strike-slip fault, and R2 is the bend featuring oblique normal slip components at relatively shallow depth. The bend R2 connects R1 and another possible strike-slip fault segment R3, and is not well mapped due to inadequate seismicity in that area. The three fault segments R1, R2, and R3 are later used as receiver faults in Coulomb stress transfer calculation. We show that during the reactivation, R1 is mainly characterized by right lateral strike-slip faults and forms an extensional bend at its western end, which is consistent with the observed oblique normal slip.

We also map the strike-slip fault system from seismicity in cross-section view (Figure 4.1f). At depth (> 7.7km), a strike-slip fault is mapped as the main structure, while at shallower depth, multiple splays from the main strike-slip fault form a flower structure. Slip within the flower structure comprises of a mix of strike-slip and normal slip components. Similar flower structures in Oklahoma have been observed in seismic reflection profiles, for example, Figure

S4.1 shows a flower structure mapped on a fault in Osage County (northeast Oklahoma). However, detailed seismic reflection profile is not available in the study area.

4.4 Fault Reactivation Process

To study the fault reactivation process, we examine the temporal distribution of earthquakes in the Woodward cluster. As shown in Figure 4.2, several spikes appear in the daily seismicity rate curve (blue line), and steps in the cumulative moment curve (red line). Each spike or step results from an increase in the number of events. We separate the sequence into 6 stages based on temporal clustering of the events with a threshold of 1.5 of the daily seismicity rate based on magnitude completeness of 2.5 using the C. Chen (2016) catalog. If several peaks in the seismicity rate curve are close in time and the corresponding events are also close in space, those peaks are then manually grouped into one stage. As shown in Figure 4.3, the earthquakes during each stage (red dots) are highly clustered in space and represent a newly reactivated fault segment. The seismicity starts at a shallower depth (stage 1) and then migrates to the northeast to a deeper depth (stage 2). From stage 3 to stage 6, we observe a bilateral migration pattern of the earthquakes. One direction is to the northeast at a relatively deep depth ($\sim$ 7.5km), which is still on the main fault segment R1; the other direction is to the southwest at a shallower depth ($\sim$ 6km), which corresponds to fault segments R2 and R3.

During each stage, the earthquakes are closely clustered in both space and time, so we speculate that earthquake interaction has a significant influence on the evolution of the cluster. This mechanism has been considered for

induced seismicity (e.g., Brown & Ge, 2018; X. Chen et al., 2017; Schoenball & Ellsworth, 2017a; Sumy et al., 2014). To test this hypothesis, we calculate the Coulomb stress change (Lin & Stein, 2004; Toda et al., 2005) from the earthquakes in each stage on the three receiver fault segments (R1, R2, and R3) using the Coulomb3.3 program (Toda et al., 2011) with friction coefficient of 0.68. The value is chosen as an average of experimental data on several rock types sampled from OK basement, at experimental depth conditions of 6km. The data is consistent with hydrothermal granite friction from Blanpied et al. (1991) and (Kolawole et al., 2018). We parameterize each stage as a source fault as follows: we calculate a cumulative magnitude from the cumulative moment of the highly clustered events in each stage and take the magnitude-weighted average location of the events as the location of the source fault. The geometries (strike, dip, and rake) of each source fault are the average of the dominant focal mechanism solutions in the subcluster. The information used to calculate the source faults is listed in Table S1. The source fault geometries are listed in Table 4.1b. The source faults represent active patches of the existing fault system (R1-R3), and the parameters in Table 4.1b are the input to calculate fault width, length, and slip using empirical relations in Coulomb3.3 program (Toda et al., 2011). Stage 1 is at relatively shallow depth and featured by normal slip component, and stage 2 is characterized by a right lateral strike-slip fault. From stage 3 to stage 6, the bilateral migration is characterized by different faulting types in northeast (NE) direction and southwest (SW) direction. We use two source faults to represent the bilateral migration in each stage from 3 to 6. The NE migration section is along the main strike-slip fault. The SW migration is mainly characterized by normal faulting on the fault bend R2. Some events to the further SW direction show

a strike-slip faulting (stage 4-SW), which might represent the rupture on R3. The fault segments in stage 3-SW and 5-NE are not mapped because no focal mechanism solutions are available. The source faults are then used to compute the Coulomb stress transfer on the receiver faults R1-R3. The results are shown in Figure 4.4.

By overlaying the earthquake locations from next stage to Coulomb stress transfer map (Figure 4.4), we find that the rupture front follows the pattern of positive Coulomb stress transfer. The cluster is initiated during stage 1 within the shallow bend of the fault comprising of dominant normal slip components. This slip generates positive Coulomb stress on the main fault (R1) to the northeast, where the next rupture front in stage 2 occurs. Then the events in stage 2 generate positive Coulomb stress on both fault bend (R2) and on the main fault further to the northeast (R1), which matches the bilateral seismicity migration in the following stages. The northeast migration is characterized by strike-slip faulting, which implies the extension of the main strike-slip fault. The southwest migration is at shallower depth and mainly reactivates the normal faulting structures of R2. The Coulomb stress transfer from individual stages 1,2, and 3 is as high as 0.15MPa and those stages dominate the rupture pattern of the fault system. A complete focal mechanism catalog (15% of events have available focal mechanisms) is not yet available for this study, so we cannot calculate Coulomb stress transfer on each earthquake. Instead, we count the percentage of earthquakes from next stage that fall within the areas of positive Coulomb transfer by interpolating the gridded stress map (listed in Table 4.1c). The source events in stage 1 and 2 are more dispersed possibly due to fewer stations at the beginning of the cluster. To test the influence of source fault location, we randomly pick two other locations in the earthquake

clouds, and the locations show a significant influence on the results (Figure S4.3). The choice of magnitude-weighted average location for source faults in our analysis is more reasonable, considering that larger events may have higher location accuracy and tend to generate larger Coulomb stress changes. The variations of friction coefficient are tested with values of $0.2, 0.4, 0.6, 0.8$ and 1.0 in Figure S4.4. Different friction coefficients do not significantly influence the results. The median value difference between $\mu = 1.0$ and $\mu = 0.2$ is less than 0.002MPa for all stages.

4.5 Discussion: Why is the Earthquake Interaction Important?

4.5.1 Characteristics in Seismicity Distribution vs the Poisson Process

In the viewpoint of stress interactions, the interevent time distribution of induced seismicity is different from those driven by static or dynamic stress changes, e.g., aftershock sequence. If one cluster is induced by fluid injection with no stress interaction among earthquakes, even with an elevated rate, those events are independent from each other. The cluster then still can be characterized by a Poisson process. The parameter coefficient of variation $c_v(\tau) = \sigma_\tau/\overline{\tau}$ with the standard deviation σ and the mean $\overline{\tau}$ can be used to check if one point-process is a Poisson process. The coefficient of variation of 1 is for Poisson process and the values higher than 1 are for temporally clustered occurrence times (Kagan & Jackson, 1991). For the Woodward cluster, the

c_v is 2.4, that is, the cluster is not a Poisson process and instead some stress interactions between individual events are expected, which is consistent with Schoenball and Ellsworth (2017a).

The nearest-neighbor approach provides a metric of distance by unifying the time, location, and magnitude distance between any two events (Zaliapin et al., 2008). The 2D plot of rescaled time and space distance in this approach was used to differentiate induced and natural seismicity in the Coso Geothermal Field (Schoenball et al., 2015) – induced seismicity features a large population of Poissonian background seismicity with relatively large rescaled time and space distance compared to cluster mode for natural earthquakes. We follow this approach to calculate the magnitude-scaled time and space distance between events and plot them in 2D histogram (Figure S4.5a). Most of the events are in the cluster mode below the diagonal $\log RT = 2.5$ with small scaled time and space distance. This cluster mode is well separated from the background mode by replacing the event times with random times from a Poisson process (Figure S4.5b). We use the same parameters as in Schoenball and Ellsworth (2017a) to make a direct comparison to the distribution of all seismicity in Oklahoma. Similar to their results, a large fraction of events is distributed in the cluster mode rather than the background mode, which suggests that the seismicity in Oklahoma is a complex process and driven by multiple factors besides pore pressure changes.

4.5.2 Pore Pressure, Poroelastic Stress, and Coulomb Stress Transfer

Most injection wells in this region are distributed over 40km to the east of the cluster, with one exception $\sim$ 10km to the south. Without other major tectonic activities known in that cluster, our preferred explanation is that the cluster is initiated by fluid injection, though this determination is not the purpose of this paper. The detailed pore pressure change and poroelastic stress change from fluid injection for the Woodward cluster were studied by Goebel et al. (2017). Their results showed that the pore pressure changes decreased rapidly with distance and the poroelastic stress changes became dominant at distance larger than 20km. For the Woodward cluster, the median pore pressure is 0.03MPa, and the median poroelastic stress change is about 0.12MPa. The Coulomb stress change from earthquakes for the first three individual stages can readily reach 0.15MPa. The median values for all grids on the faults are around 0.02MPa with the dip angle used in this analysis (Figure S4.4). However, we find that the dip angle has a strong influence on the stress amplitude. If we instead use steep dip angle (i.e., near vertical) calculated from seismicity cloud, the median values easily reach 0.05 to 0.06MPa during later stages (Figure S4.6). The results from the steeper dip angle are more comparable to the 2D faults assumed in Goebel et al. (2017). The comparison suggests that Coulomb stress changes from earthquakes are at least comparable to pore pressure and poroelastic stress changes in the Woodward cluster from wastewater disposal. Therefore, while fluid injection may have caused the initial rupture at the fault bend, the internal Coulomb stress interaction among different fault patches at least assists the occurrence of the subsequent earthquake sequence. We

should mention that the pore pressure change from earthquakes is considered in Coulomb stress calculation via Skempton's coefficient and the trace element of the stress tensor (e.g., Sumy et al., 2014).

Previous studies also examined cumulative Coulomb stress changes from earthquakes and suggested that the Coulomb stress transfer can have strong impact on the subsequent earthquake locations (e.g., Brown & Ge, 2018; Catalli et al., 2013; Pennington & Chen, 2017). The cumulative Coulomb stress transfer in our study reaches 0.3MPa. This level of stress change is significant enough to influence the evolution of the earthquake sequence (e.g., Catalli et al., 2013; King et al., 1994; Pennington & Chen, 2017; Rydelek & Sacks, 1999). The earthquake interaction can be an important triggering mechanism and should be considered for comprehensive hazard assessment for induced seismicity.

4.5.3 Role of Fault Zone Structure in Fault Activation Process

The Woodward cluster reveals a complex flower structure with a mixture of normal and strike-slip faulting. The seismicity during stage 1 initiated at shallow depth with dominant normal slip, which caused a positive Coulomb stress transfer in the area that ruptured during stage 2. Extensional "jogs" or releasing "bends" can be relatively weak and typically experience "loading-weakening" during the inter-seismic period (Sibson, 1993). Relatively lower strength is more likely to be triggered by external stress perturbations. Previous studies have observed dynamic triggering in volcanic or geothermal fields where higher fluid content has been perceived (e.g., Aiken & Peng, 2014; Hill, 2006; Hough & Kanamori, 2002). Foreshocks leading to large earthquakes also tend

to occur within extensional jogs (X. Chen & Shearer, 2013). Detailed analysis of an earthquake sequence in central Oklahoma suggested the sequence may have initiated within an extensional jog at shallower depth (X. Chen et al., 2018). The foreshocks of M5.8 Pawnee earthquake occurred along a conjugate fault system and resulted in a combined increase in stress of up to 0.7MPa at the mainshock hypocenter (X. Chen et al., 2017; Pennington & Chen, 2017). These studies, along with the observations in this study suggest that fault zone structure has a strong influence on the initiation and evolution of earthquake sequences, whether induced or natural.

4.6 Conclusions

We mapped fault structures in Woodward using accurately relocated seismicity. The reactivated fault is a strike-slip fault system, with the main strike-slip fault at depth and a flower structure with oblique and normal slip components above. The observed structures are consistent with the transtensional stress field in northwest Oklahoma and fault geometries imaged in 3D seismic data in northeast Oklahoma. The seismicity distribution shows that the mapped faults are reactivated step by step, and the rupture pattern can be explained at least in part by Coulomb stress interaction from earthquakes. Also, the cluster characteristics show evidences of earthquake interactions. By comparing the calculated Coulomb stress transfer to pore pressure and poroelastic stress changes, we conclude that the cluster might be initiated by fluid injection, but the stress interaction is an important factor in the ongoing reactivation of fault segments.

Acknowledgments

This work was supported by Oklahoma Governor's Emergency Funding for Induced Seismicity and a student fellowship at the ConocoPhillips School of Geology and Geophysics at the University of Oklahoma. The two earthquake relocation files used in this study can be found in the supporting material and at `https://pubs.geoscienceworld.org/ssa/srl/article/88/5/1252/354106/waveform-relocated-earthquake-catalog-for-oklahoma`, respectively. The focal mechanism solution file can be found in the supporting file. We thank Dr. Chris Rollins and another anonymous reviewer for their valuable comments and suggestions.

References

Aiken, C., & Peng, Z. (2014). Dynamic triggering of microearthquakes in three geothermal/volcanic regions of California. *Journal of Geophysical Research: Solid Earth*, *119*(9), 6992-7009. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JB011218 doi: 10.1002/2014JB011218

Blanpied, M. L., Lockner, D. A., & Byerlee, J. D. (1991). Fault stability inferred from granite sliding experiments at hydrothermal conditions. *Geophysical Research Letters*, *18*(4), 609-612. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/91GL00469 doi: 10.1029/91GL00469

Brown, M. R. M., & Ge, S. (2018). Small earthquakes matter in injection-induced seismicity. *Geophysical Research Letters*, *45*(11), 5445-5453. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL077472 doi: 10.1029/2018GL077472

Catalli, F., Meier, M.-A., & Wiemer, S. (2013). The role of Coulomb stress changes for injection-induced seismicity: The Basel enhanced geothermal system. *Geophysical Research Letters*, *40*(1), 72-77. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012GL054147 doi: 10.1029/2012GL054147

Chen, C. (2016). *Comprehensive analysis of Oklahoma earthquakes: From earthquake monitoring to 3d tomography and relocation* (Unpublished doctoral dissertation). University of Oklahoma.

Chen, X., Jackson, H., W., G. T. H., Xiaofeng, M., Zhigang, P., & C., C. J. (2018). Temporal correlation between seismic moment and injection volume for an induced earthquake sequence in central Oklahoma. *Journal of Geophysical Research: Solid Earth*, *123*(4), 3047-3064. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JB014694 doi: 10.1002/2017JB014694

Chen, X., Nakata, N., Pennington, C., Haffener, J., Chang, J. C., He, X., ... Walter, J. I. (2017). The Pawnee earthquake as a result of the interplay among injection, faults and foreshocks. *Scientific Reports*, *7*(1), 1–18. Retrieved from http://dx.doi.org/10.1038/s41598-017-04992-z doi: 10.1038/s41598-017-04992-z

Chen, X., & Shearer, P. M. (2013). California foreshock sequences suggest aseismic triggering process. *Geophysical Research Letters*, *40*(11), 2602-2607. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/grl.50444 doi: 10.1002/grl.50444

Ellsworth, W. L. (2013). Injection-induced earthquakes. *Science*, *341*(6142). Retrieved from http://science.sciencemag.org/content/341/6142/1225942 doi: 10.1126/science.1225942

Goebel, T. H., Weingarten, M., Chen, X., Haffener, J., & Brodsky, E. E. (2017). The 2016 Mw5.1 Fairview, Oklahoma earthquakes: Evidence for long-range poroelastic triggering at >40 km from fluid disposal wells. *Earth and Planetary Science Letters*, *472*, 50–61. Retrieved from http://dx.doi.org/10.1016/j.epsl.2017.05.011 doi: 10.1016/j.epsl.2017.05.011

Hardebeck, J. L., Survey, U. S. G., & Park, M. (2005). HASH : A FORTRAN program for computing earthquake first-motion focal mechanisms – v1.1 – August 17 , 2005. *Computing*, 1–15.

Harding, T. (1985). Seismic characteristics and identification of negative flower structures, positive flower structures, and positive structural inversion. *American Association of Petroleum Geologists Bulletin*, *69*(4), 582–600. doi: 10.1306/AD462538-16F7-11D7-8645000102C1865D

Hill, D. P. (2006). Unrest in Long Valley Caldera, California, 1978–2004. *Geological Society, London, Special Publications*, *269*(1), 1–24. Retrieved from `http://sp.lyellcollection.org/content/269/1/1.2` doi: 10 .1144/GSL.SP.2006.269.01.02

Hough, S. E., & Kanamori, H. (2002). Source properties of earthquakes near the salton sea triggered by the 16 October 1999 M 7.1 Hector Mine, California, Earthquake. *Bulletin of the Seismological Society of America*, *92*(4), 1281. Retrieved from `http://dx.doi.org/10.1785/0120000910` doi: 10.1785/0120000910

Kagan, Y. Y. (2005). Double-couple earthquake focal mechanism: random rotation and display. *Geophysical Journal International*, *163*(3), 1065–1072.

Kagan, Y. Y., & Jackson, D. D. (1991). Long-term earthquake clustering. *Geophysical Journal International*, *104*(1), 117-133. Retrieved from `http://dx.doi.org/10.1111/j.1365-246X.1991.tb02498.x` doi: 10 .1111/j.1365-246X.1991.tb02498.x

Keranen, K. M., Savage, H. M., Abers, G. A., & Cochran, E. S. (2013). Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 Mw 5.7 earthquake sequence. *Geology*, *41*(6),

699–702. doi: 10.1130/G34045.1

King, G. C. P., Stein, R. S., & Lin, J. (1994). Static stress changes and the triggering of earthquakes. *Bulletin of the Seismological Society of America*, *84*(3), 935. Retrieved from http://dx.doi.org/

Kolawole, F., Johnston, C. S., Chang, J. C., Morgan, C. B., Marfurt, K., Lockner, D. A., ... Carpenter, B. M. (2018). Brittle deformation in the precambrian igneous basement of oklahoma..

Langenbruch, C., Dinske, C., & Shapiro, S. A. (2011). Inter event times of fluid induced earthquakes suggest their Poisson nature. *Geophysical Research Letters*, *38*(21), 1–6. doi: 10.1029/2011GL049474

Le Guerroué, E., & Cobbold, P. R. (2006). Influence of erosion and sedimentation on strike-slip fault systems: insights from analogue models. *Journal of Structural Geology*, *28*(3), 421 - 430. Retrieved from http://www.sciencedirect.com/science/article/pii/S0191814105002142 doi: 10.1016/j.jsg.2005.11.007

Lin, J., & Stein, R. S. (2004). Stress triggering in thrust and subduction earthquakes and stress interaction between the southern San Andreas and nearby thrust and strike-slip faults. *Journal of Geophysical Research: Solid Earth*, *109*(B2), 1–19. Retrieved from http://doi.wiley.com/10.1029/2003JB002607 doi: 10.1029/2003JB002607

McGarr, A. (2014). Maximum magnitude earthquakes induced by fluid injection. *Journal of Geophysical Research: Solid Earth*, *119*(2), 1008–1019. doi: 10.1002/2013JB010597

Pennington, C., & Chen, X. (2017). Coulomb stress interactions during the Mw 5.8 Pawnee Sequence. *Seismological Research Letters*, *88*(4), 1024. Retrieved from http://dx.doi.org/10.1785/0220170011 doi:

10.1785/0220170011

Qin, Y. (2017). *Study of stress field in seismogenic processes in oklahoma* (Unpublished master's thesis). University of Oklahoma.

Reasenberg, P., & Oppenheimer, D. (1985). FPFIT, FPPLOT, and FPPAGE: FORTRAN computer programs for calculating and displaying fault-plane solutions. *U.S. Geol. Surv. Open File Report*(85), 85–739.

Rydelek, P. A., & Sacks, I. S. (1999). Large earthquake occurrence affected by small stress changes. *Bulletin of the Seismological Society of America*, *89*(3), 822. Retrieved from http://dx.doi.org/

Schoenball, M., Davatzes, N. C., & Glen, J. M. G. (2015). Differentiating induced and natural seismicity using space-time-magnitude statistics applied to the Coso Geothermal field. , 1–8. doi: 10.1002/2015GL064772

Schoenball, M., & Ellsworth, W. L. (2017a). A systematic assessment of the spatiotemporal evolution of fault activation through induced seismicity in Oklahoma and southern Kansas. *Journal of Geophysical Research: Solid Earth*, 189–206. doi: 10.1002/2017JB014850

Schoenball, M., & Ellsworth, W. L. (2017b). Waveform-relocated earthquake catalog for Oklahoma and southern Kansas illuminates the regional fault network. *Seismological Research Letters*. doi: 10.1785/0220170083

Segall, P., & Lu, S. (2015). Injection-induced seismicity: Poroelastic and earthquake nucleation effects. *Journal of Geophysical Research: Solid Earth*, *120*(7), 5082-5103. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JB012060 doi: 10.1002/2015JB012060

Sibson, R. H. (1993). Load-strengthening versus load-weakening faulting. *Journal of Structural Geology*, *15*(2), 123 - 128. Retrieved from http://www

.sciencedirect.com/science/article/pii/019181419390090W doi: 10.1016/0191-8141(93)90090-W

Sumy, D. F., Cochran, E. S., Keranen, K. M., Wei, M., & Abers, G. A. (2014). Observations of static Coulomb stress triggering of the November 2011 M5. 7 Oklahoma earthquake sequence. *Journal of Geophysical Research: Solid Earth*, *119*(3), 1904–1923. doi: 10.1002/2013JB010612

Sylvester, A. (1988). Strike-slip faults. *Geological Society of America Bulletin*(November), 1–38. Retrieved from http://gsabulletin.gsapubs.org/content/100/11/1666.short doi: 10.1130/0016-7606(1988)100$\langle$1666:SSF$\rangle$2.3.CO;2

Toda, S., Stein, R. S., Richards-Dinger, K., & Bozkurt, S. B. (2005). Forecasting the evolution of seismicity in southern California: Animations built on earthquake stress transfer. *Journal of Geophysical Research: Solid Earth*, *110*(5), 1–17. doi: 10.1029/2004JB003415

Toda, S., Stein, R. S., Sevilgen, V., & Lin, J. (2011). Coulomb 3.3 Graphic-rich deformation & stress-change software for earthquake, tectonic and volcano research and teaching - User Guide. *USGS Open-File Report 2011-1060*, 63. Retrieved from https://pubs.usgs.gov/of/2011/1060/of2011-1060.pdf{%}0Ahttp://pubs.usgs.gov/of/2011/1060/

Vidale, J. E., & Shearer, P. M. (2006). A survey of 71 earthquake bursts across southern California: Exploring the role of pore fluid pressure fluctuations and aseismic slip as drivers. *Journal of Geophysical Research: Solid Earth*, *111*(5), 1–12. doi: 10.1029/2005JB004034

Waldhauser, F., & Ellsworth, W. L. (2000). A Double-difference earthquake location algorithm: Method and application to the Northern Hayward Fault, California. *Bulletin of the Seismological Society of America*, *90*(6),

1353–1368. doi: 10.1785/0120000006

Windows, F. O. R. (2009). Seisan: the Earthquake Analysis Software. (May), 8952.

Yeck, W. L., Hayes, G. P., McNamara, D. E., Rubinstein, J. L., Barnhart, W. D., Earle, P. S., & Benz, H. M. (2017). Oklahoma experiences largest earthquake during ongoing regional wastewater injection hazard mitigation efforts. *Geophysical Research Letters*, *44*(2), 711–717. doi: 10.1002/2016GL071685

Zaliapin, I., & Ben-Zion, Y. (2015). Artefacts of earthquake location errors and short-term incompleteness on seismicity clusters in southern California. *Geophysical Journal International*, *202*(3), 1949–1968. doi: 10.1093/gji/ggv259

Zaliapin, I., Gabrielov, A., Keilis-Borok, V., & Wong, H. (2008). Clustering analysis of seismicity and aftershock identification. *Physical Review Letters*, *101*(1), 4–7. doi: 10.1103/PhysRevLett.101.018501

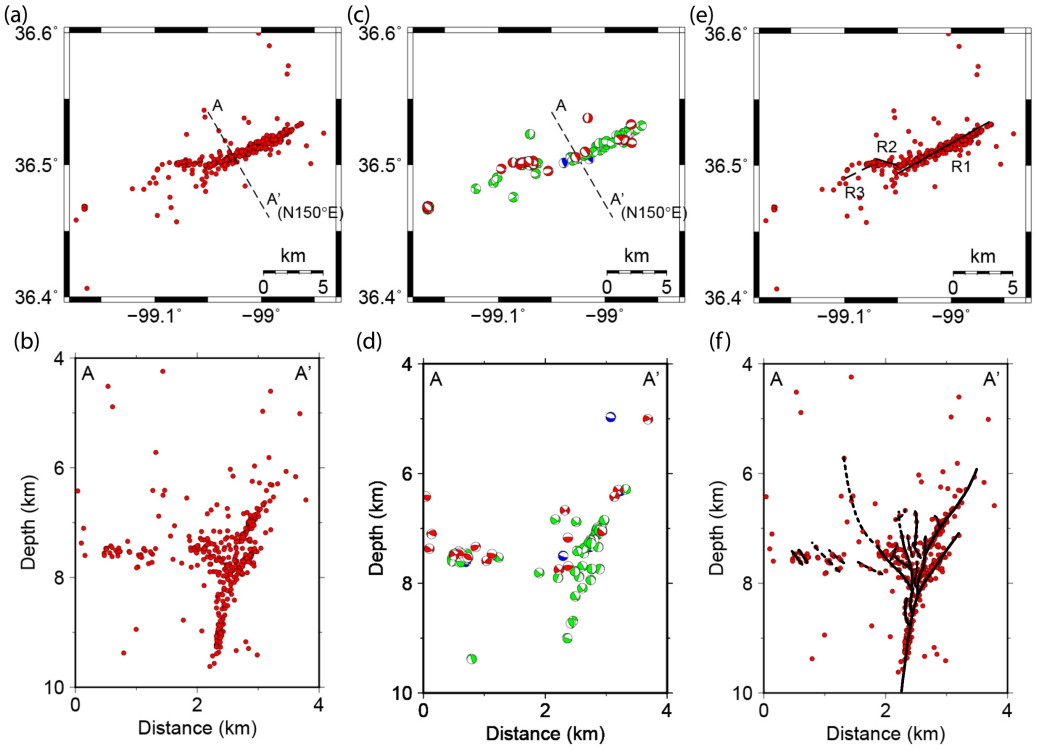

Figure 4.1: Seismicity, focal mechanisms and fault interpretation in the Woodward cluster. (a) Seismicity distribution in map view. (b) Seismicity distribution in AA′ cross-section view. (c) Focal mechanism distribution in map view using lower hemisphere projection. (d) Focal mechanism distribution in AA′ cross-section view. Red: normal faulting; green: strike-slip faulting; blue: reverse faulting in (c) and (d). (e) Interpreted fault segments (black lines) in map view. (f) Interpreted fault segments (black lines) in AA′ cross-section view. Dashed black lines represent segments where fault geometry is uncertain in (e) and (f). Relocations are from Schoenball and Ellsworth (2017b).

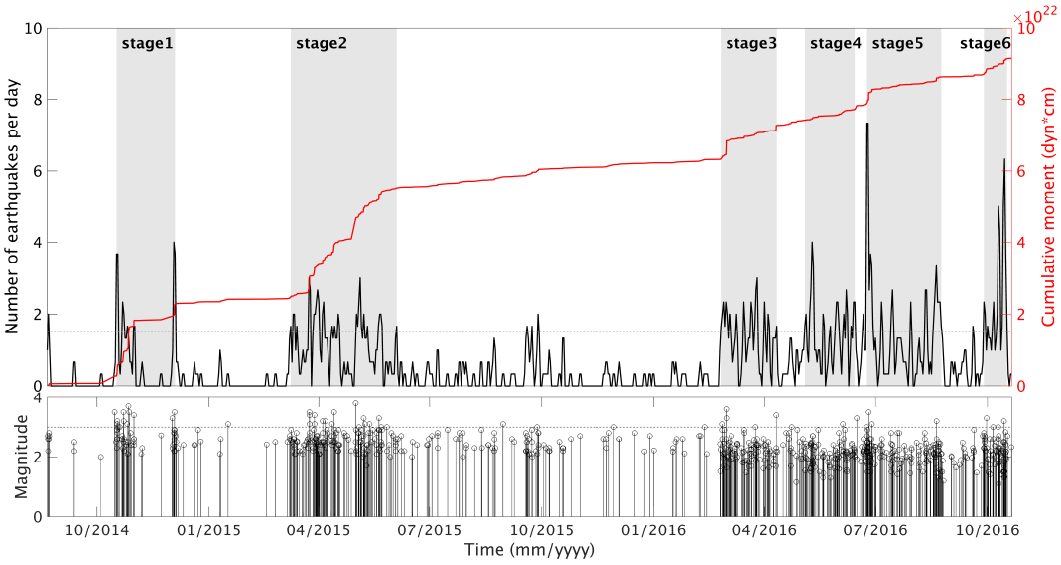

Figure 4.2: Time distribution of seismicity in the Woodward cluster. Top: daily seismicity rate (black) and cumulative moment (red) curves. The dashed line denotes the daily seismicity rate of 1.5. Six stages are highlighted based on the date ranges with daily seismicity rate larger than 1.5. Bottom: magnitude distribution along time. The dashed line denotes the magnitude of 3.0. Note that the maximum magnitude in this cluster is smaller than 4.0. Relocations are from C. Chen (2016).

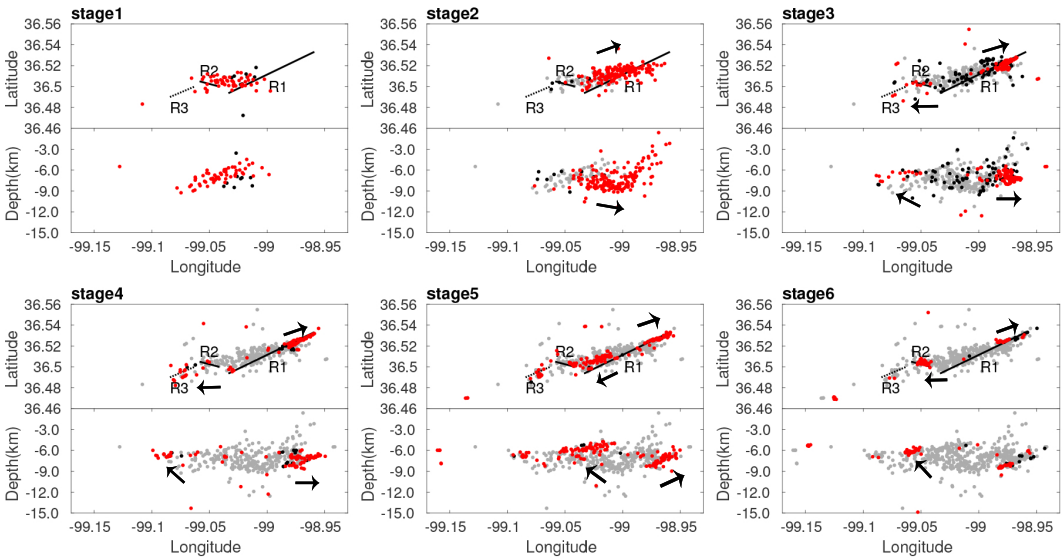

Figure 4.3: Spatial distribution of seismicity in the Woodward cluster. For each stage, the top panel shows the map distribution and the bottom panel shows the depth distribution. Red: events in current stage. Gray: events in previous stages. Black: events between last stage and current stage. The receiver faults mapped in Figure 1 are shown under the seismicity. The seismicity migration is denoted by black arrows. Note the bilateral migration of seismicity to the northeast (deeper) and southwest (shallower) in the cluster. Relocations are from C. Chen (2016).

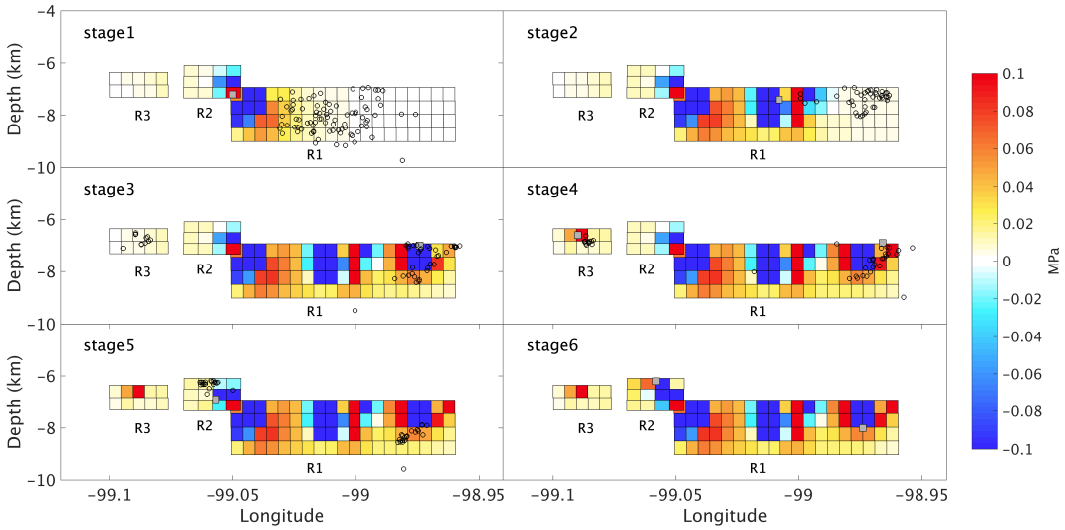

Figure 4.4: Coulomb stress transfer calculated from the source faults (Table 4.1b) on the receiver faults (Table 4.1a), μ=0.68. Each panel shows the cumulative Coulomb stress transfer from stage 1 to current stage. The open circles are earthquake locations from next stage. The gray squares represent the source fault patches in each stage.

Table 4.1: Fault Parameters Mapped from Seismicity and Focal Mechanism Solutions.

(a) Receiver Fault Parameters

Fault	Start point	End point	Top (km)	Bottom (km)	Strike (°)	Dip (°)
R1	(-99.05, 36.49)	(-98.96, 36.53)	6.93	9.01	65.6	65
R2	(-99.07, 36.51)	(-99.05, 36.50)	6.09	7.34	96.7	46
R3	(-99.10, 36.49)	(-99.08, 36.50)	6.36	7.32	56.9	80

(b) Source Fault Parameters

Stage	Mag	Lon	Lat	Depth (km)	Strike (°)	Dip (°)	Rake (°)	Type	Fault Segment
1	4.1	-99.05	36.50	7.2	58	50	-120	NF	R2
2	4.3	-99.01	36.51	7.4	63	68	-157	SS	R1
3-NE	3.9	-98.97	36.52	7.0	60	66	-155	SS	R1
4-NE	3.3	-98.97	36.52	6.8	55	64	-157	SS	R1
4-SW	2.8	-99.09	36.50	6.6	77	75	-153	SS	R3
5-SW	3.7	-99.06	36.51	6.9	168	25	-107	NF	R2
6-NE	3.0	-98.97	36.53	8.0	56	74	158	SS	R1
6-SW	3.6	-99.06	36.50	6.2	107	47	-92	NF	R2

SS: strike-slip fault; NF: normal fault.

(c) Earthquake Triggering Statistics

Stage	Total number of EQs from next stage	EQs in ΔCFS>0	EQs in ΔCFS>0.01MPa
1	89	81	19
2	42	39	2
3	47	19	19
4	39	25	25
5	43	23	23
6	-	-	-

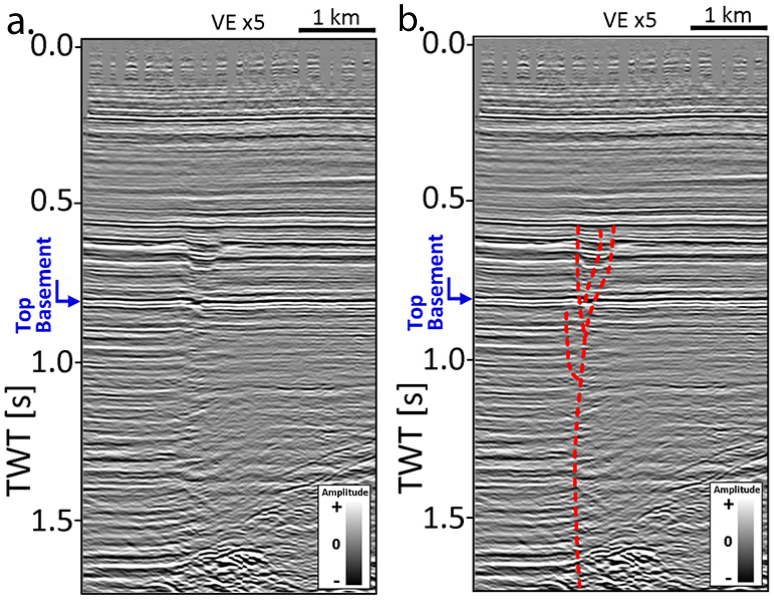

Figure S4.1: Example of a flower structure mapped from a seismic reflection profile. (a) Seismic amplitude cross-section from a 3D seismic reflection data in Osage county, Oklahoma. (b) Interpreted flower structure (red dotted line) from seismic reflection data profile in Figure S4.1a. The seismic data is a post-stack time-migrated 3D volume provided by SpyGlass Energy.

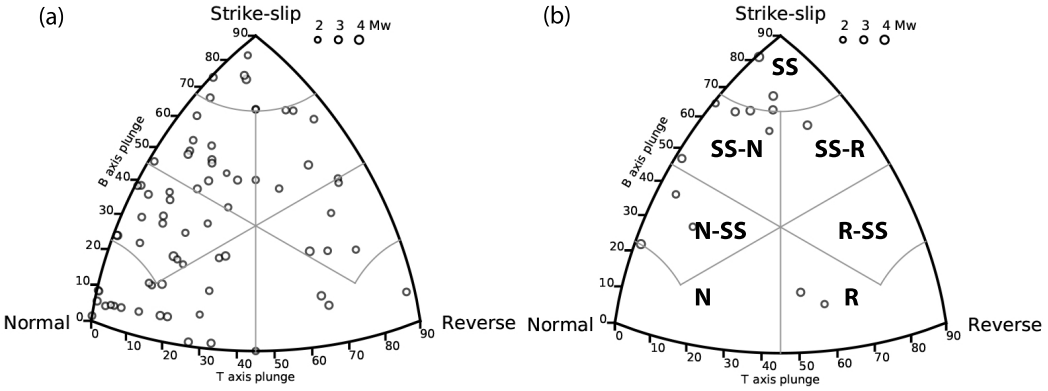

Figure S4.2: Focal mechanisms classification diagrams for (a) FMS shallower than 7.7km and (b) FMS deeper than 7.7km. The classification diagram is based on SMT axes plunges by Kagan (2005). SS: strike-slip faulting, R: reverse faulting, and N: normal faulting. The depth of 7.7km separates the deeper strike-slip faults from shallower normal faults.

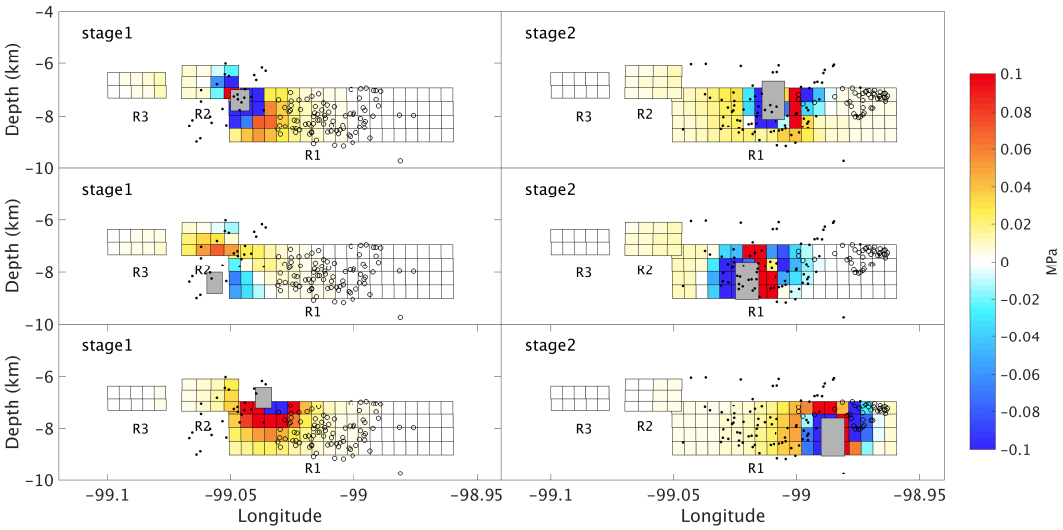

Figure S4.3: Coulomb stress transfer calculated from the source faults 1 and 2 with different locations. The first row shows the magnitude-weighted average locations used in this paper. The second and third row show two randomly selected locations in the corresponding earthquake cloud.

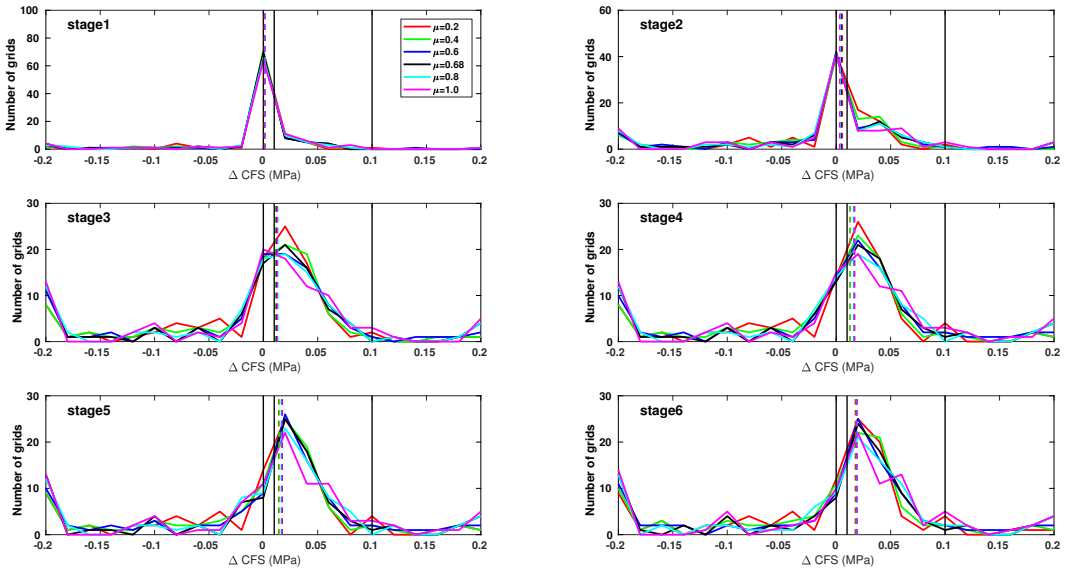

Figure S4.4: Histogram of Coulomb stress transfer of all grids on receiver faults (Table 1a) under different friction coefficients. The vertical, solid lines denote the Coulomb stress transfer of 0, 0.01MPa and 0.1MPa. The vertical, dashed lines denote the median value of ΔCFS of all grids on receiver faults.

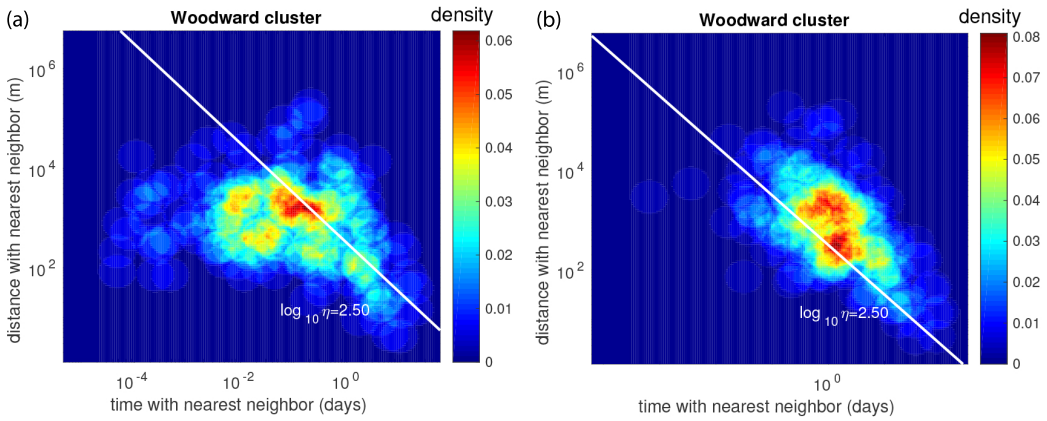

Figure S4.5: 2D density plot of nearest distance T and R for (a) events in the Woodward cluster and (b) events in the cluster with event times replaced by random times from a Poisson process.

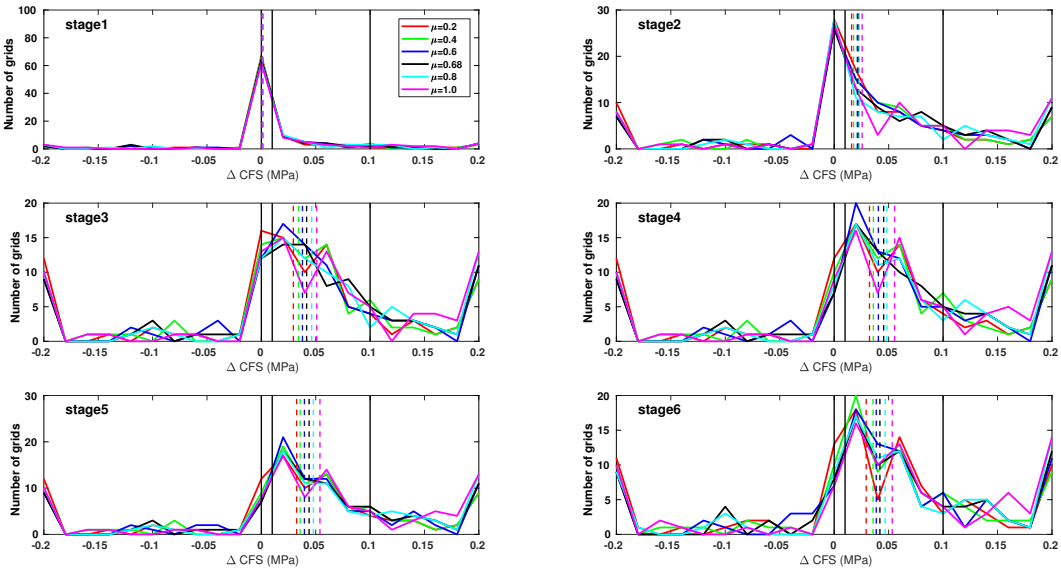

Figure S4.6: Histogram of Coulomb stress transfer of all grids on receiver faults (Table 1a with strikes of 88°, 80°, and 87° from earthquake cloud) under different friction coefficients. The vertical, solid lines denote the Coulomb stress transfer of 0, 0.01MPa and 0.1MPa. The vertical, dashed lines denote the median value of ΔCFS of all grids on receiver faults.

Table S4.1: Information Used in the Calculation of Source Fault.

Stage	# events	Max Mag	Min Mag	Max dist(km)	# FMs
1	29	2.5	3.7	1.6	2
2	92	2.5	3.9	3.4	6
3-ENE	17	2.5	3.6	1.0	5
4-ENE	6	2.5	3.1	1.4	5
4-WSW	1	2.8	2.8	0	2
5-WSW	2	2.5	3.5	1.9	2
6-ENE	4	2.5	2.7	0.6	6
6-WSW	10	2.5	3.3	0.7	12

Chapter 5

Forecasting Induced Seismicity in Oklahoma using Machine Learning Methods

Plain Language Summary

The increase of earthquakes in Oklahoma in the last decade has been associated with wastewater disposal. Researchers have forecast the induced seismicity rate change using statistics- or physics-based models. In this study, we choose a machine learning method, Random Forest (RF), to directly relate injection parameters to seismicity rate change. The injection parameters, including injection rate, injection volume, injection depth, modeled pore pressure, and poroelastic stress, are used as input in the RF model to forecast the seismicity rate. The model can forecast the rapid decrease in seismicity in recent years. The model also identifies pore pressure as the most important feature

in the forecasting. The findings are consistent with the known mechanisms of induced seismicity in Oklahoma.

5.1 Introduction

The sharp seismicity increase in the last decade has been associated with wastewater disposal in Oklahoma (e.g. Ellsworth, 2013; Keranen et al., 2014; Yeck et al., 2016). The mechanisms of the induced earthquakes include pore pressure diffusion (e.g., Keranen et al., 2014; Haffener et al., 2018; Langenbruch et al., 2018; Zhai et al., 2019), poroelastic stress disturbance at a larger distance (e.g., Segall & Lu, 2015; Goebel et al., 2017), and earthquake nucleation in event clusters (e.g. Segall & Lu, 2015; Chen et al., 2017; Pennington & Chen, 2017; Qin et al., 2018). Many studies have made seismicity rate forecasts based on known mechanisms. For example, Langenbruch et al. (2018) used a hybrid physical-statistical model combining hydrogeologic modeling and modified Gutenberg-Richter relation to forecast seismic hazard in space and time. The study found that earthquake probabilities in Oklahoma and Kansas were increasing with the square of the pressure rate. Zhai et al. (2019) used a poroelastic modeling and rate-and-state earthquake nucleation model to forecast the timing and magnitude of induced seismicity. They found that pore-pressure diffusion controlled the induced earthquakes in Oklahoma and its impact was enhanced by poroelastic effects. Norbeck and Rubinstein (2018) developed a numerical model that integrated fluid pressurization from injection with a rate-and-state friction description of the earthquake nucleation process to forecast rates of induced seismicity. The paper found that the models that were informed with injection data outperform a standard statistical model

that used prior earthquake observations to forecast future induced earthquake activity.

Apart from the modeled pore pressure and poroelastic stress, some injection parameters have been directly associated with the induced seismicity. Injection rate is found to be one of the most important parameters based on various methods, including earthquake association with injections and physical modeling from injections (e.g., Barbour et al., 2017; Chen et al., 2018; Langenbruch & Zoback, 2016; Weingarten et al., 2015). Hincks et al. (2018) reported that injection depth relative to the crystalline basement most strongly correlated with seismic moment release and the joint effects of depth and volume were critical from an advanced Bayesian network. Rock properties and regional geological structures also influence the occurrence of the induced seismicity. For example, Shah and Keller (2017) showed that most earthquakes in Oklahoma were located where the crystalline basement was likely composed of fractured intrusive or metamorphic rock. Pei et al. (2018) conducted 2-D Pg wave tomography and found that most moderate-size (M>4) earthquakes occurred either close to the boundaries between high- and low-seismic velocity zones or within the high-velocity zones, suggesting the influence of geologic structures on earthquake spatial locations. The most correlated parameters with earthquakes could change with different scales of studies and different methods (e.g., Weingarten et al., 2015).

In this paper, we compile all the available physical and operational parameters from wastewater injection and input them in a Random Forest regression model to forecast the seismicity rate. Without any prior knowledge of the weight of the input features, the model can identify the most important features

and improve the understanding of the mechanisms of induced seismicity.

5.2 Data

5.2.1 Earthquakes

We use the earthquake catalog from January 2010 to February 2020 from the Oklahoma Geological Survey (OGS) (Walter et al., 2020) (Figure 5.1). Studies have shown the influence of earthquake interactions in clusters in Oklahoma (e.g., Segall & Lu, 2015; Chen et al., 2017; Pennington & Chen, 2017; Qin et al., 2018). Since the earthquake nucleation is not accounted for in this study, we decluster the catalog by removing the aftershocks following the spatial and temporal windows proposed for Oklahoma earthquakes in Rosson et al. (2019). The catalog has a completeness of magnitude (M_c) of 2.2. In this study, we focus on events with a magnitude larger than 2.7 due to the M_c change over time.

5.2.2 Wastewater Injection

We download the monthly injection data from 2010 to 2018 from the Oklahoma Corporation Commission (OCC). We select wells with total injection depth larger than 1000 ft and the cumulative injection volume since 2010 larger than 1000 bbl (Figure 5.1). We also interpolate the basement measurements in Campbell (2003) at each injection well location to calculate the distance between the injection well and the basement. The final operational parameters include monthly injection rate, yearly and cumulative injection volume, injection depth, basement depth, and distance from the well bottom to the basement. Since

the injection data are currently not available for the year 2019 and 2020 on OCC website, we assume the injection parameters stay constant that is equal to the average of the last three reported months of injection data in 2018.

5.2.3 Pore Pressure and Poroelastic Stress Data

To include physics-based features from injections, we add pore pressure data (Figure 5.1) and poroelastic stress data from hydrogeologic modeling. The pore pressure data are from Langenbruch et al. (2018). The data were calculated for each month at 25000 randomly selected points in Oklahoma and have been used in a statistical model to successfully forecast statewide seismic hazard in Oklahoma. The poroelastic data are from Zhai et al. (2019), which were calculated at gridded points in central and western Oklahoma by assuming an optimal NW and NE fault orientation. The data have been used in seismicity forecast in a pore pressure based rate-and-state model. The pore pressure data (Langenbruch et al., 2018) and poroelastic stress data (Zhai et al., 2019) are derived from two hydrogeologic models with different parameters. The pore pressure data cover the whole study area and have better spatial and temporal resolution. There is likely inconsistency of assumed modeling parameters between the two studies.

5.3 Methods

We select the study area from -99.5° to -96° in longitude and from 35° to 37° in latitude (inset figure in Figure 5.1), which covers the majority of the seismicity in Oklahoma. The study area is divided into uniform grid cells,

and in each cell, we search for earthquakes, injection wells, and modeled pore pressure and poroelastic stress data points (Figure 5.1) for each month. The statistics, including sum, mean, maximum, minimum, different percentiles, variance, skewness, and kurtosis, of the injection parameters, pore pressures, and poroelastic stress data in that grid in the current month are used as input features. The number of earthquakes in the next time window is the target to predict. We should mention that we extend the grid cell (referred to as grid plus) to search for wells, pore pressure, and poroelastic points because some earthquakes at the grid borders might be associated with the injection wells outside the grid cell. We choose different grid sizes from 0.4° to 1.0° with an increment of 0.1° and prediction windows of 30, 60, 90, and 180 days to build the input features and the target. We also apply a moving window of one-third of the grid size to get more data points. The range of grid sizes and window sizes allows us to test the spatial and temporal resolution of the forecasting model.

We use a machine learning technique Random Forest (RF) to forecast the monthly seismicity rate from the input features. RF model constructs a multitude of decision trees based on the features in training datasets and outputs the mean prediction of the individual trees for the regression problem (Ho, 1995, 1998). In a decision tree, the model selects one feature to best split the data into two groups, and in each group, the model repeats the process to build different levels of the tree. The details of the model structure are described in Rouet-Leduc et al. (2017). We use the python package Scikit-learn (Pedregosa et al., 2011) to implement the RF model.

We split the data into training (2010–2016) and test (2017–2020) dataset.

The trained model should be able to learn the complete relationships between seismicity rate and the features since the training data include the onset, increase, peak, and part of the decrease of the seismicity rate. Evidence has shown that the seismicity rate changes over time (e.g., Llenos & Michael, 2013; Montoya-Noguera & Wang, 2017), so the earthquake forecasting is a time-dependent series. To account for the temporal changes in the model, we add the difference by comparing features in current month to the previous month as new features. We also have features from previous month, the average of previous 2–5 months, and the average of previous 6–10 months as short-, medium-, and long-term history data, respectively.

The input data are then normalized based on the training data set, and the highly correlated features are removed. We select 50 most important features and perform a grid search of the hyperparameters. The best-fit model is used to make predictions for the test data. A metric of r^2 score is selected to measure how well the regression predictions approximate the real data points. The r^2 score is defined as,

$$r^2 = \frac{\sum_i (f_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} \tag{5.1}$$

where $\bar{y}$ is the mean of observed data y_i, f_i is the corresponding prediction of y_i. An r^2 of 1 indicates that the regression predictions perfectly fit the data. A constant model that always predicts the expected value of y, disregarding the input features, would result in an r^2 score of 0.0.

The model also outputs the relative importance of the feature to the predictability of the target variable. This predictive power of the feature is estimated by the fraction of samples a feature contributes to combined with

the decrease in impurity from splitting them (Pedregosa et al., 2011). The rank values are between 0 and 1, with values close to 1 meaning the highest importance. The rank of the features helps understand the physics behind the induced seismicity. We have different statistics in different windows as features. To make direct comparison, we sum the feature importance of all related statistics in different time windows together for each category of feature.

5.4 Results

For input from each combination of grid sizes and prediction time windows, we search for the best-fit RF model using training data and get the predictions for the test data from the best-fit model. The results are discussed in the next section. Here, we show the results from the grid size of 0.9° and the prediction window of 30 days, which have a relatively high r^2 score of 0.48 for the test dataset. The number of training and test data points is 4752 and 2508, respectively. The best-fit model has the hyperparameters as follows, tree number of 10, maximum depth of 5, the minimum number of samples to split of 4, the minimum number of samples for a leaf node of 16, and the maximum number of features to split of 10. The parameters are relatively small compared to other studies (e.g., Rouet-Leduc et al., 2017), to prevent overfitting on a small dataset.

The map view of forecasting results is shown in Figure 5.2. The model predicts the number of earthquakes in the time window of 30 days, and the annual seismicity prediction is the sum of the earthquake number in each month of a year. Consistent with the observations, the seismicity is mainly distributed in central and western Oklahoma, and the seismicity rate remains elevated

until 2020 especially for western Oklahoma. The drastic color change between 2017 and 2018 suggests that the model captures the rapid decrease of seismicity in those two years. Since we apply a moving window to the grid input, the predictions are smoothed out, and no abrupt local changes are predicted by the model.

We plot the temporal seismicity curve for different regions in Figure 5.3 by summing the predictions in individual grids. The model predicts the decreasing trend of seismicity after 2016 well compared to the true observations. The seismicity level in 2020 remains high as that in 2013. In subregions of central and western Oklahoma, the predicted curve and the observations fit well. For local clusters, e.g., the three local M$\geq$5 sequences in Pawnee, Prague, and Fairview, the forecasting results are consistent with the true observations. The results suggest that our model can forecast the seismicity rate with a spatial resolution of 0.9 °. We also notice that the forecast in 2019 and 2020 shows slight over-prediction, which could result from the assumption of constant injection rate after 2018.

For comparison, we use a linear regression method to forecast seismicity based on the same set of features, and the results are shown in Figure S5.1 and S5.2. The linear model predicts the decreasing trend of seismicity, but for some regions, it results in large discrepancy between the observed and predicted data. However, the linear model show similar relative importance of the features as the RF model.

Figure 5.4 shows the rank of feature importance from the model. The most important features are pore pressure, poroelastic stress, injection rate, injection volume, and injection depth. For each feature category, all the related

feature importance is summed together. The pore pressure and poroelastic stress correspond to the two main mechanisms of induced seismicity. Since the modeling of pore pressure and poroelastic stress incorporates the injection rate, injection depth, and other hydrogeologic parameters, the modeled physics-based features show higher importance than the individual operational parameters. Besides the physics-based features, the injection rate and injection depth also show high importance, consistent with the previous studies (e.g., Langenbruch & Zoback, 2016; Hincks et al., 2018).

The physics-based parameters or the operational parameters alone can also make reasonable forecasting of the seismicity (Figure S5.3, S5.4 and Figure S5.5, S5.6). The data from the physics-based model show slightly better and more smooth results than those from model based only on operational parameters, probably because the effect of abrupt changes in injection rate has been diluted by the diffusion process. By checking all the individual features in each category, the most important pore pressure feature is the most recent mean pore pressure rate, and the most important poroelastic stress feature is the 20th percentile of the squared poroelastic stress. For injection volume, the most recent features show larger importance than medium- and long-term history data. Other than that, no unambiguous conclusion could be drawn about the relative importance of recent history and long-term history for other features.

5.5 Discussion

5.5.1 Parameters to Build Input: Grid Size, Prediction Window, Grid Plus

Different scales of spatial-temporal relations between fluid injection and earthquakes have been found in Oklahoma, from local clusters (e.g., Walsh & Zoback, 2015; Chen et al., 2017) to regional scales (e.g., Yeck et al., 2016; Pollyea et al., 2018). Haffener et al. (2018) found robust spatial correlations between injection and seismicity at a range of different grid sizes and various time delays with different grid sizes. We test on a range of grid sizes (0.4° to 1.0° with an increment of 0.1°) and prediction time windows (30, 60, 90, and 180 days). For each combination of spatial and temporal windows, we search for the best-fit RF model. The results are shown in Figure S5.7 and S5.8. The goodness of the fit (r^2 score) is relatively high for the grid size of 0.7 to 0.9°. The grid input for smaller grid size probably overlooks some regional correlations between earthquakes and injections (e.g., Yeck et al., 2016). As for the prediction windows, the model makes better predictions for short-term windows. The model tends to overpredict the seismicity rate for longer windows, which is probably due to the rapid decrease of input features in test data.

To account for the correlations between earthquakes and injection close to borders, we extend the grid to search for responsible wells for the earthquakes in the grid. Figure S5.9 shows the r^2 score variations for different grid plus sizes. Grid plus of 0.1° and 0.2° generates better results than a grid plus size of 0 by including the injection wells associated with the earthquakes at the grid borders. The r^2 score starts to decrease if the grid plus is too large (e.g., 0.3°),

which suggests that at a larger distance (around 30km), the wells should not be associated with the earthquakes within the grid. The results are consistent with the spatial range of pore pressure diffusion influence from previous studies (e.g., Keranen et al., 2014; Yeck et al., 2016).

5.5.2 Local Stress State and Spatial Heterogeneity

Besides the selected features in this study, the local geological features and stress state also influence the relationship between injection and earthquakes. Shah and Keller (2017) found that most earthquakes were located likely in fractured intrusive or metamorphic crystalline basement and Pei et al. (2018) found that most moderate-size earthquakes occurred either close to the boundaries between high- and low-seismic velocity zones or within the high-velocity zones, suggesting that they were associated with geological boundaries of different basement rock properties or with strong material properties in the upper crust. Hincks et al. (2018) found that a geospatial correction parameter was highly correlated with the annual seismic release.

In our model, to characterize the spatial heterogeneity, we also add columns to represent different regions. We assign each grid a specific region number and encode the K categorical region numbers to K region columns with zeros and ones. For each entry in the input, only the column related to its region number is set to one, and all other region columns are set to zero. The prediction results by adding the features do not show significant improvement, and the feature importance of region columns is low (<0.001). One explanation is that we select a relatively large grid size of 0.9°, which generates the best fitting compared to other grid sizes, and the local heterogeneity is averaged out. A

smaller grid size results in a larger importance for region parameter (Figure S5.10), but it is still negligible compared to other features. Also, we apply an overlapping window to build the input, which results in larger data points but could possibly smooth out the spatial heterogeneity effect.

5.5.3 Potential Magnitude Forecast

The maximum magnitude of induced earthquakes from injections is essential to evaluate seismic hazard. In previous studies on earthquake forecasting, the forecast seismicity rate is transferred to the probability of medium to large earthquake causing damage by the Gutenberg-Richter (GR) frequency-magnitude relationship. For example, Zhai et al. (2019) forecast the seismicity rate using a rate-and-state model and estimated the occurrence probability of M5 earthquakes. In this study, we only forecast the time and location of earthquakes using temporal and spatial prediction windows. The model obtains similar results to those from traditional, statistics- or physics-based methods. Next, we could also assume a GR frequency-magnitude relationship and calculate the probability of moderate to large earthquakes. In the machine learning model, we could also add a target variable of the maximum magnitude to estimate the seismic hazards. The maximum magnitude from injection has been related to the total injection volume (McGarr, 2014) or seismogenic index (Van der Elst et al., 2016), which is selected features or could be added to the feature list in this study. So it is highly possible to get the maximum magnitude estimation and a complete earthquake forecast from our machine learning model.

5.6 Conclusions

In this paper, we compile the injection-related parameters and directly relate the parameters to the seismicity rate in a Random Forest regression model. The model can predict the induced seismicity rate as precisely as traditional methods. The seismicity is mainly distributed in central and western Oklahoma, and the seismicity rate will remain elevated in 2020. The model also ranks the features by their importance, which implies the mechanisms behind induced seismicity. Pore pressure, poroelastic stress, injection rate, and injection depth are the most important features. The model demonstrates its potential in earthquake forecasting and understanding the physics behind the induced seismicity without making many assumptions.

References

Barbour, A. J., Norbeck, J. H., & Rubinstein, J. L. (2017). The effects of varying injection rates in osage county, oklahoma, on the 2016 m w 5.8 pawnee earthquake. *Seismological Research Letters*, *88*(4), 1040–1053.

Campbell, J. A. (2003). Wells drilled to basement in oklahoma.

Chen, X., Haffener, J., Goebel, T. H., Meng, X., Peng, Z., & Chang, J. C. (2018). Temporal correlation between seismic moment and injection volume for an induced earthquake sequence in central oklahoma. *Journal of Geophysical Research: Solid Earth*, *123*(4), 3047–3064.

Chen, X., Nakata, N., Pennington, C., Haffener, J., Chang, J. C., He, X., ... Walter, J. I. (2017). The pawnee earthquake as a result of the interplay among injection, faults and foreshocks. *Scientific reports*, *7*(1), 1–18.

Ellsworth, W. L. (2013). Injection-induced earthquakes. *Science*, *341*(6142), 1225942. doi: 10.1126/science.1225942

Goebel, T., Weingarten, M., Chen, X., Haffener, J., & Brodsky, E. (2017). The 2016 mw5. 1 fairview, oklahoma earthquakes: Evidence for long-range poroelastic triggering at¿ 40 km from fluid disposal wells. *Earth and Planetary Science Letters*, *472*, 50–61.

Haffener, J., Chen, X., & Murray, K. (2018). Multiscale analysis of spatiotemporal relationship between injection and seismicity in oklahoma. *Journal*

of Geophysical Research: Solid Earth, *123*(10), 8711–8731.

Hincks, T., Aspinall, W., Cooke, R., & Gernon, T. (2018). Oklahoma's induced seismicity strongly linked to wastewater injection depth. *Science*, *359*(6381), 1251–1255.

Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, pp. 278–282).

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, *20*(8), 832–844.

Keranen, K. M., Weingarten, M., Abers, G. A., Bekins, B. A., & 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

Langenbruch, C., Weingarten, M., & Zoback, M. D. (2018). Physics-based forecasting of man-made earthquake hazards in oklahoma and kansas. *Nature communications*, *9*(1), 1–10. doi: 10.1038/s41467-018-06167-4

Langenbruch, C., & Zoback, M. D. (2016). How will induced seismicity in oklahoma respond to decreased saltwater injection rates? *Science advances*, *2*(11), e1601542.

Llenos, A. L., & Michael, A. J. (2013). Modeling earthquake rate changes in oklahoma and arkansas: Possible signatures of induced seismicity. *Bulletin of the Seismological Society of America*, *103*(5), 2850–2861.

McGarr, A. (2014). Maximum magnitude earthquakes induced by fluid injection. *Journal of Geophysical Research: solid earth*, *119*(2), 1008–1019.

Montoya-Noguera, S., & Wang, Y. (2017). Bayesian identification of multiple seismic change points and varying seismic rates caused by induced

seismicity. *Geophysical Research Letters*, *44*(8), 3509–3516.

Norbeck, J., & Rubinstein, J. L. (2018). Hydromechanical earthquake nucleation model forecasts onset, peak, and falling rates of induced seismicity in oklahoma and kansas. *Geophysical Research Letters*, *45*(7), 2963–2975.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others (2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, *12*(Oct), 2825–2830.

Pei, S., Peng, Z., & Chen, X. (2018). Locations of injection-induced earthquakes in oklahoma controlled by crustal structures. *Journal of Geophysical Research: Solid Earth*, *123*(3), 2332–2344.

Pennington, C., & Chen, X. (2017). Coulomb stress interactions during the m w 5.8 pawnee sequence. *Seismological Research Letters*, *88*(4), 1024–1031.

Pollyea, R. M., Mohammadi, N., Taylor, J. E., & Chapman, M. C. (2018). Geospatial analysis of oklahoma (usa) earthquakes (2011–2016): Quantifying the limits of regional-scale earthquake mitigation measures. *Geology*, *46*(3), 215–218.

Qin, Y., Chen, X., Carpenter, B. M., & Kolawole, F. (2018). Coulomb stress transfer influences fault reactivation in areas of wastewater injection. *Geophysical Research Letters*, *45*(20), 11–059.

Rosson, Z., Walter, J., Goebel, T., & Chen, X. (2019). Narrow spatial aftershock zones for induced earthquake sequences in oklahoma. *Geophysical Research Letters*, *46*(17-18), 10358–10366. doi: 10.1029/2019GL083562

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophysical Research Letters*, *44*(18), 9276–9282.

Segall, P., & Lu, S. (2015). Injection-induced seismicity: Poroelastic and

earthquake nucleation effects. *Journal of Geophysical Research: Solid Earth*, *120*(7), 5082–5103.

Shah, A. K., & Keller, G. R. (2017). Geologic influence on induced seismicity: Constraints from potential field data in oklahoma. *Geophysical Research Letters*, *44*(1), 152–161.

Van der Elst, N. J., Page, M. T., Weiser, D. A., Goebel, T. H., & Hosseini, S. M. (2016). Induced earthquake magnitudes are as large as (statistically) expected. *Journal of Geophysical Research: Solid Earth*, *121*(6), 4575–4590.

Walsh, F. R., & Zoback, M. D. (2015). Oklahoma?s recent earthquakes and saltwater disposal. *Science advances*, *1*(5), e1500195.

Walter, J. I., Ogwari, P., Thiel, A., Ferrer, F., Woelfel, I., Chang, J. C., ... Holland, A. A. (2020). The oklahoma geological survey statewide seismic network. *Seismological Research Letters*, *91*(2A), 611–621.

Weingarten, M., Ge, S., Godt, J. W., Bekins, B. A., & Rubinstein, J. L. (2015). High-rate injection is associated with the increase in us mid-continent seismicity. *Science*, *348*(6241), 1336–1340.

Yeck, W., Weingarten, M., Benz, H. M., McNamara, D. E., Bergman, E., Herrmann, R., ... Earle, P. (2016). Far-field pressurization likely caused one of the largest injection induced earthquakes by reactivating a large pre-existing basement fault structure. *Geophysical Research Letters*, *43*(19). doi: 10.1002/2016GL070861

Zhai, G., Shirzaei, M., Manga, M., & Chen, X. (2019). Pore-pressure diffusion, enhanced by poroelastic stresses, controls induced seismicity in oklahoma. *Proceedings of the National Academy of Sciences*, *116*(33), 16228–16233.

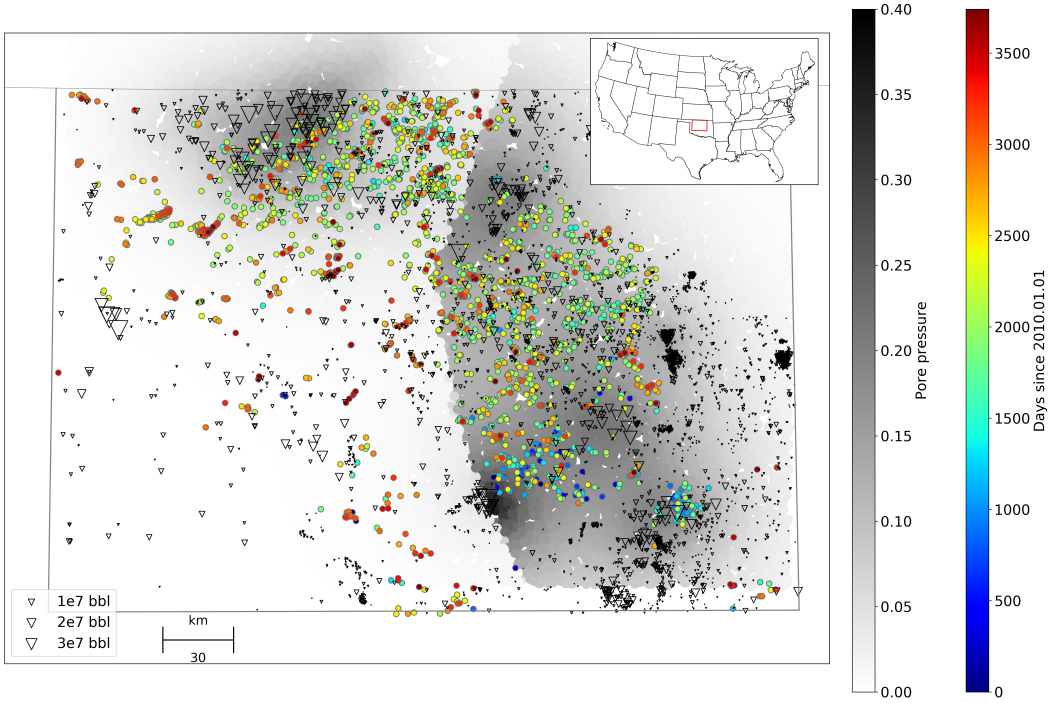

Figure 5.1: Map view of the seismicity and well injections. Triangles: injection wells scaled by the cumulative injection volume from 2010 to 2018. Circles: M>=2.7 earthquakes from 2010 to February 2020, colored by their origin times. The background shows the modeled pore pressure in June, 2015 from Langenbruch et al. (2018). The inset figure shows the location of our study area.

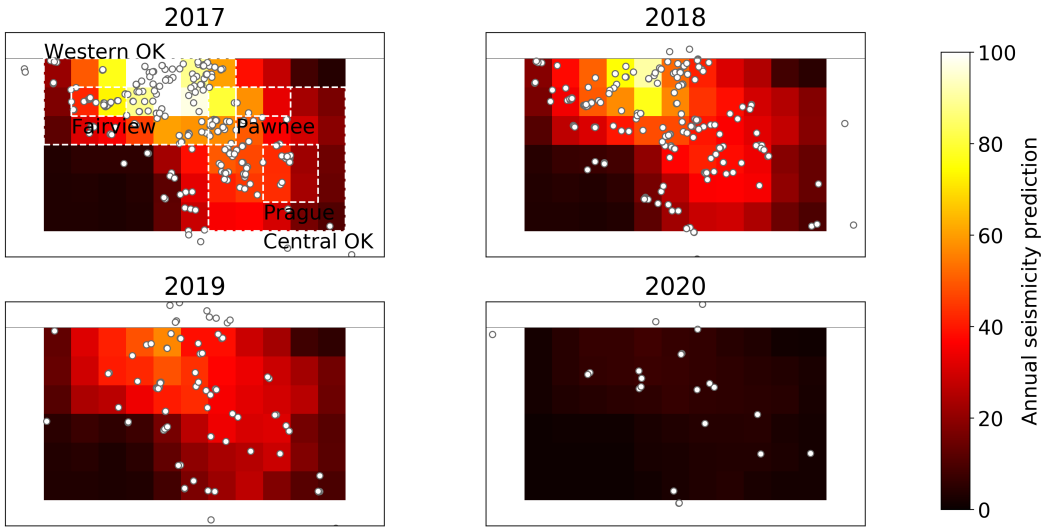

Figure 5.2: The forecast earthquake number from the model. The annual forecast is the sum of the monthly prediction in each year. The first panel shows the location of subregions used in next figure.

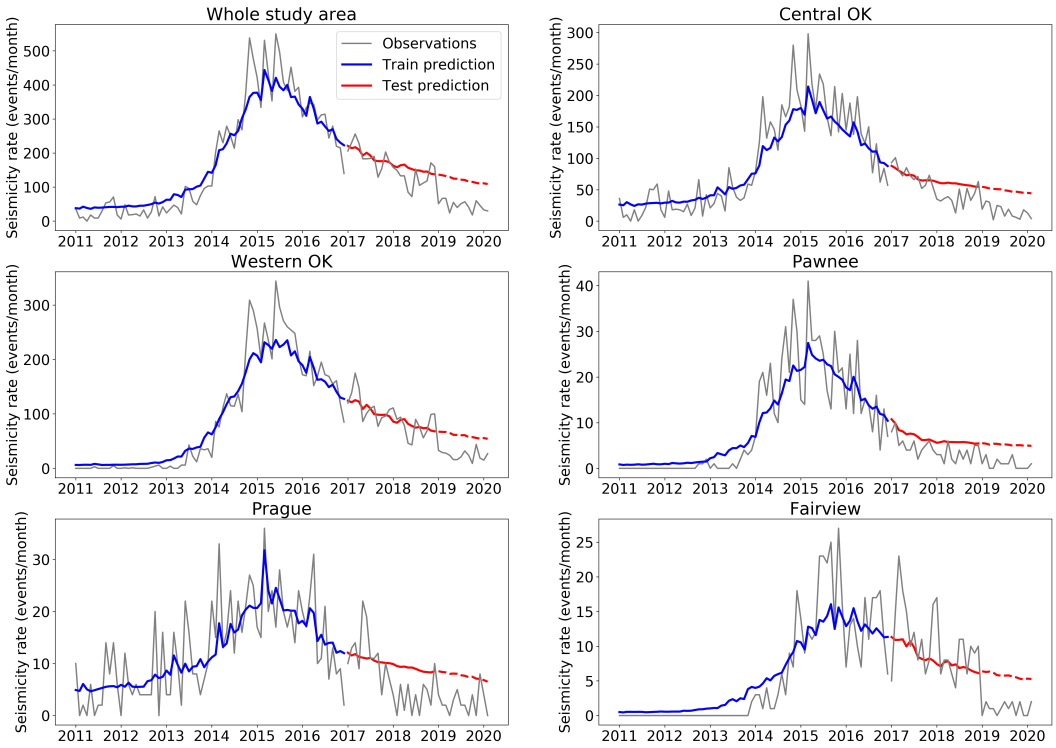

Figure 5.3: Seismicity rate forecast in (a) the whole study area, (b) central Oklahoma, (c) western Oklahoma, (d) Pawnee sequence, (e) Prague sequence, and (f) Fairview sequence. The gray lines are observations, and the blue and red lines are forecasting results from training and test dataset, respectively. The red dashed line shows forecast for year 2019 and 2020, where the injection data are not available yet. The location of the subregions is shown in Figure 5.2.

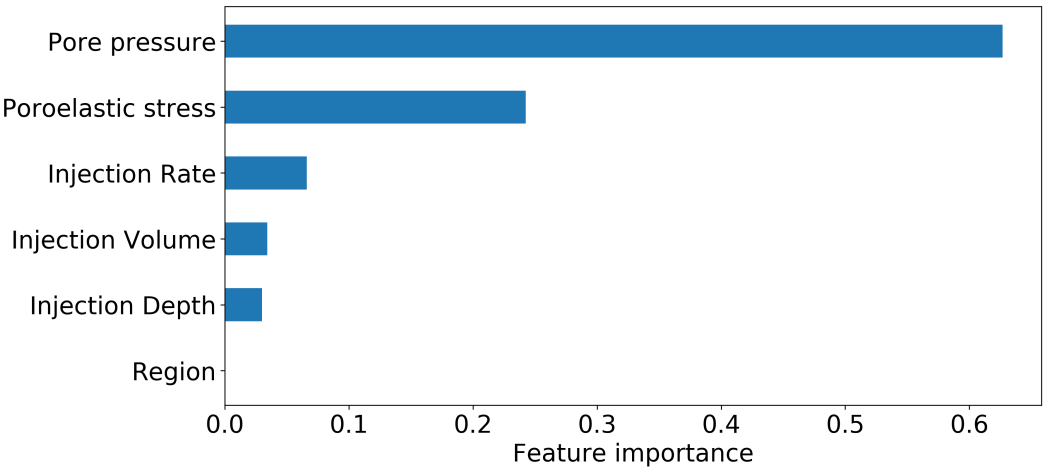

Figure 5.4: Histogram of feature importance from the model. The most important features are pore pressure, poroelastic stress, injection rate, injection volume, and injection depth. For each category of feature, the feature importance of all related features is summed together.

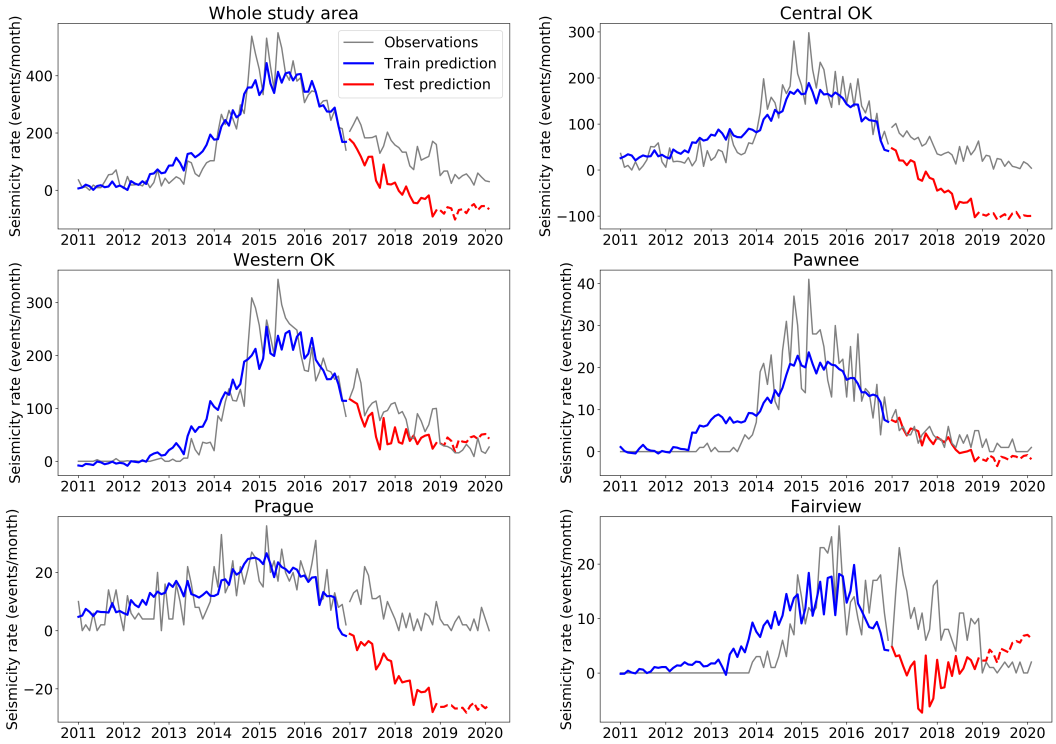

Figure S5.1: Seismicity rate forecast using linear regression in (a) the whole study area, (b) central Oklahoma, (c) western Oklahoma, (d) Pawnee sequence, (e) Prague sequence, and (f) Fairview sequence. The gray lines are observations, and the blue and red lines are forecasting results from training and test dataset, respectively. The red dashed line shows forecast for year 2019 and 2020, where the injection data are not available yet.

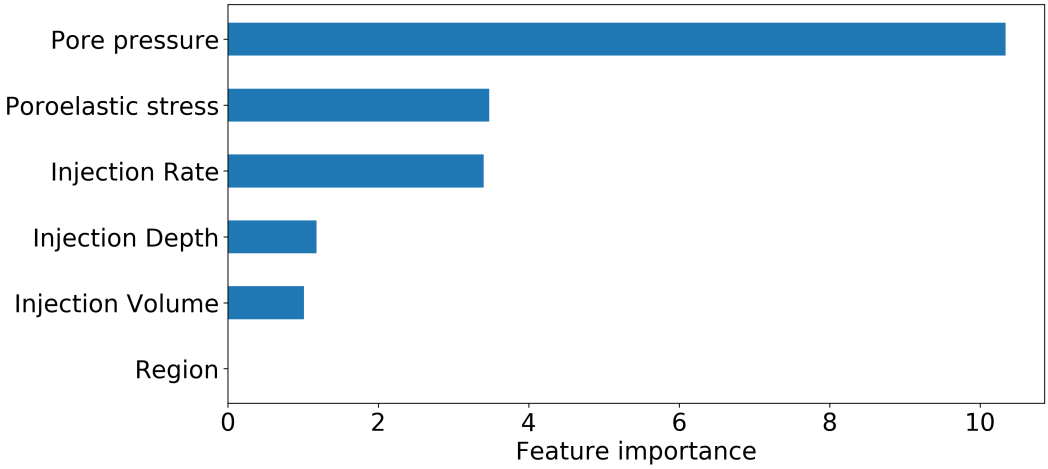

Figure S5.2: Histogram of feature importance for linear model. The feature importance are the sum of the absolute value of the coefficients of the linear model.

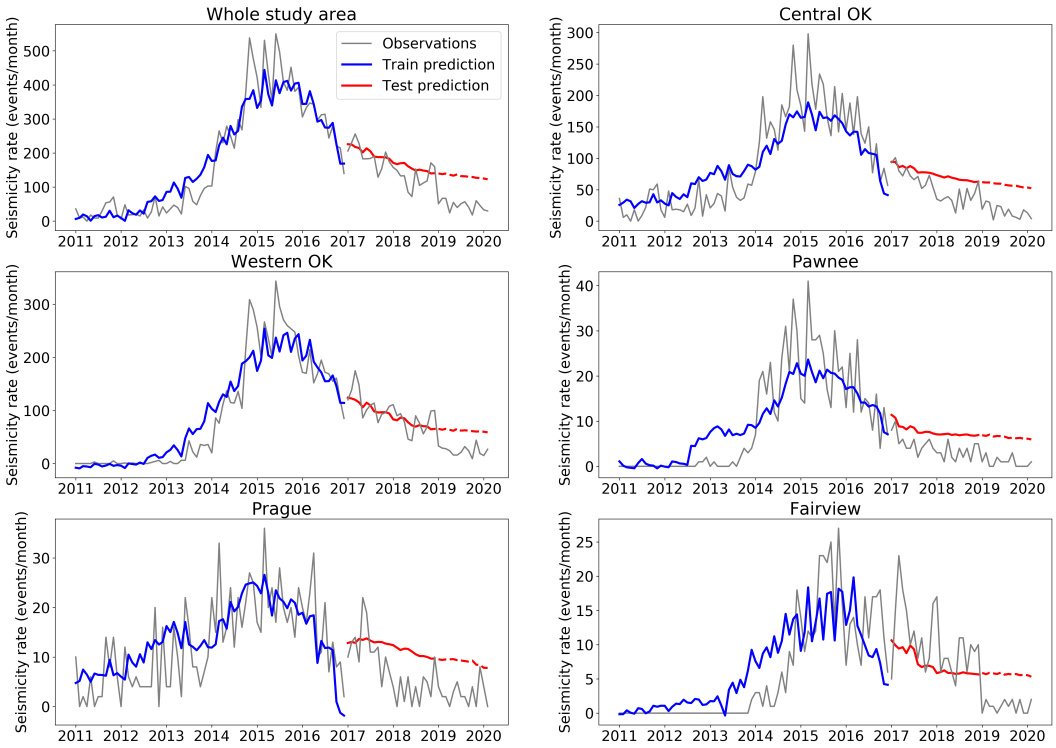

Figure S5.3: Seismicity rate forecast using only pore pressure and poroelastic stress related features in random forest in (a) the whole study area, (b) central Oklahoma, (c) western Oklahoma, (d) Pawnee sequence, (e) Prague sequence, and (f) Fairview sequence. The gray lines are observations, and the blue and red lines are forecasting results from training and test dataset, respectively. The red dashed line shows forecast for year 2019 and 2020, where the injection data are not available yet. The R^2 score is 0.38.

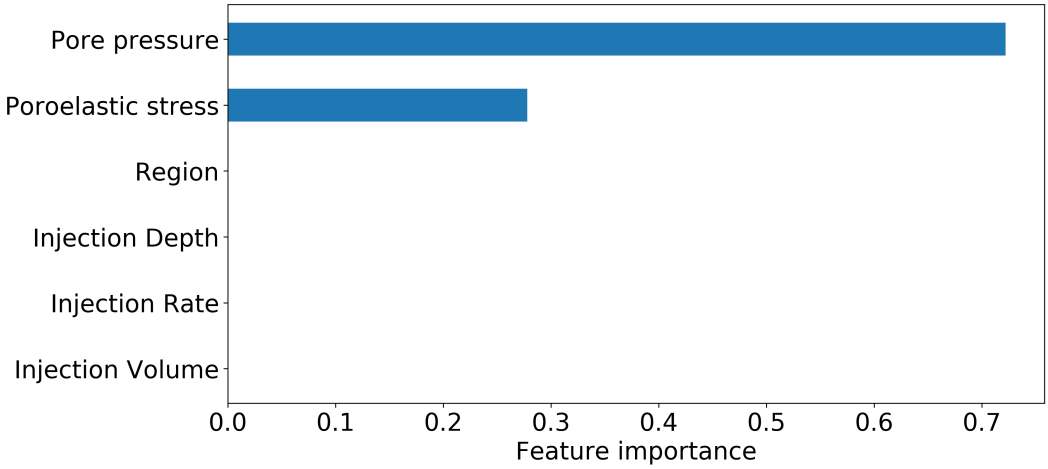

Figure S5.4: Histogram of feature importance for random forest model using only pore pressure and poroelastic stress related features. For each type of feature, the feature importance of all related features is summed together.

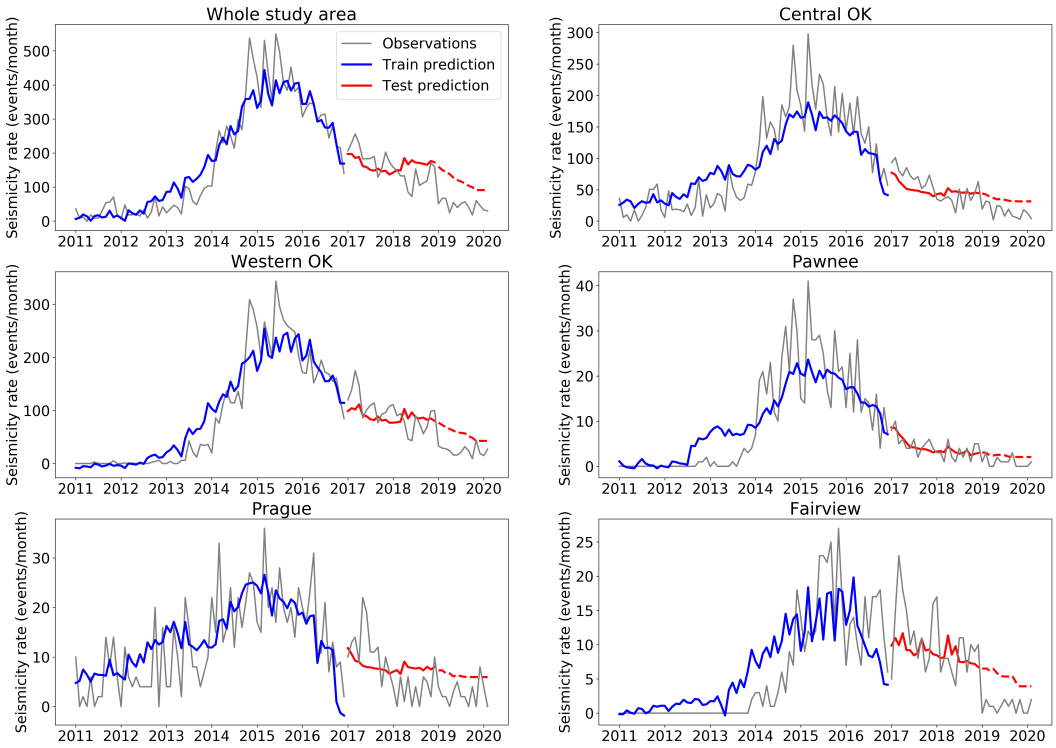

Figure S5.5: Seismicity rate forecast using only operational parameters in random forest in (a) the whole study area, (b) central Oklahoma, (c) western Oklahoma, (d) Pawnee sequence, (e) Prague sequence, and (f) Fairview sequence. The gray lines are observations, and the blue and red lines are forecasting results from training and test dataset, respectively. The red dashed line shows forecast for year 2019 and 2020, where the injection data are not available yet. The R^2 score is 0.35.

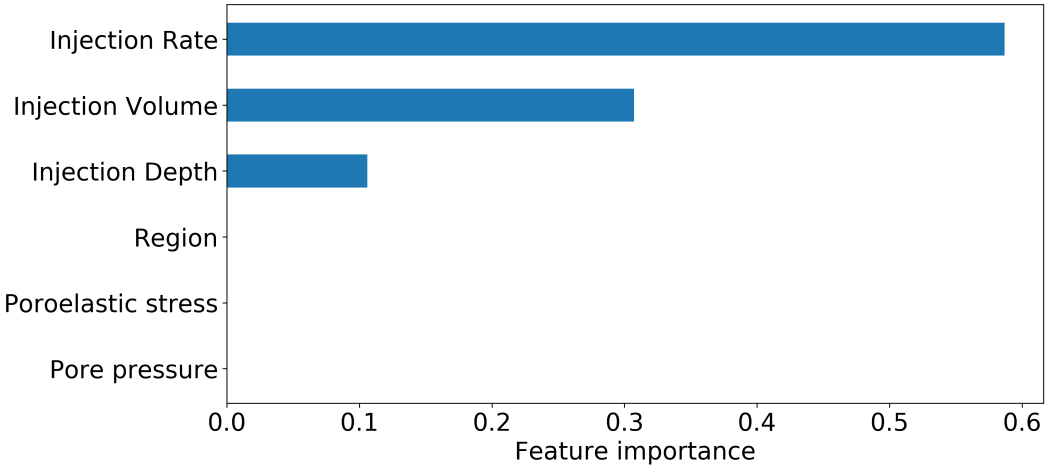

Figure S5.6: Histogram of feature importance for random forest model using only operational parameters. For each type of feature, the feature importance of all related features is summed together.

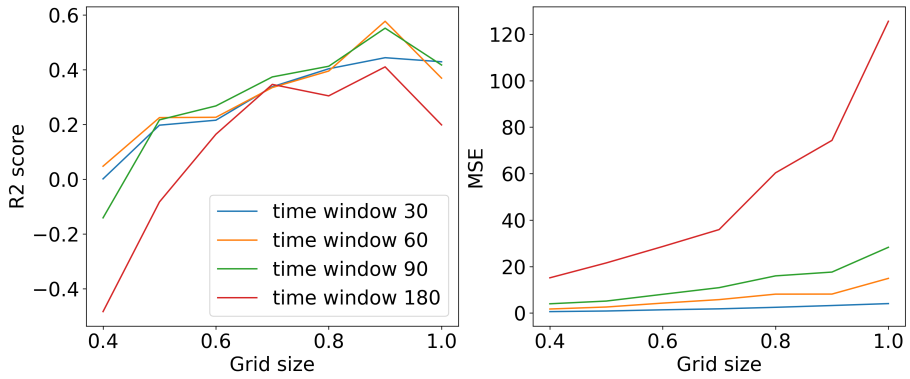

Figure S5.7: R^2 score and MSE (mean squared error) variations with grid size.

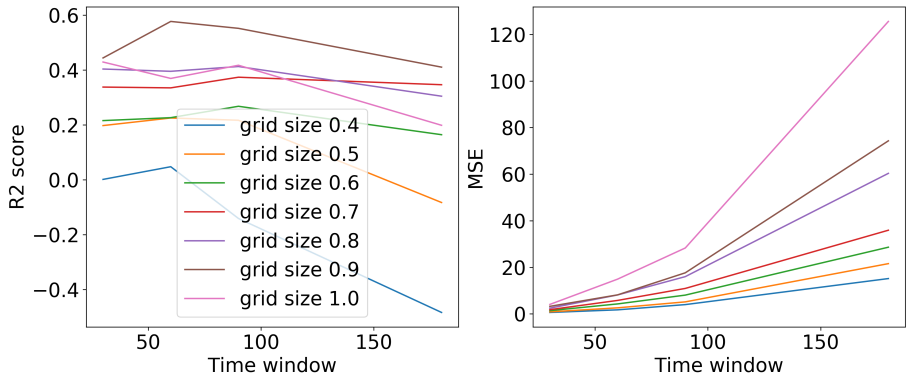

Figure S5.8: R^2 score and MSE variations with prediction window size.

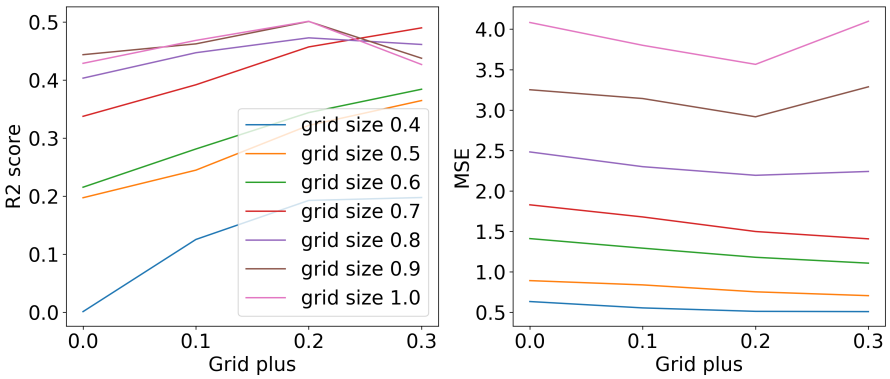

Figure S5.9: R^2 score and MSE variations with grid plus parameter. The grid plus parameter is an enlarged area around each grid to search for wells, pore pressure, and poroelastic stress points.

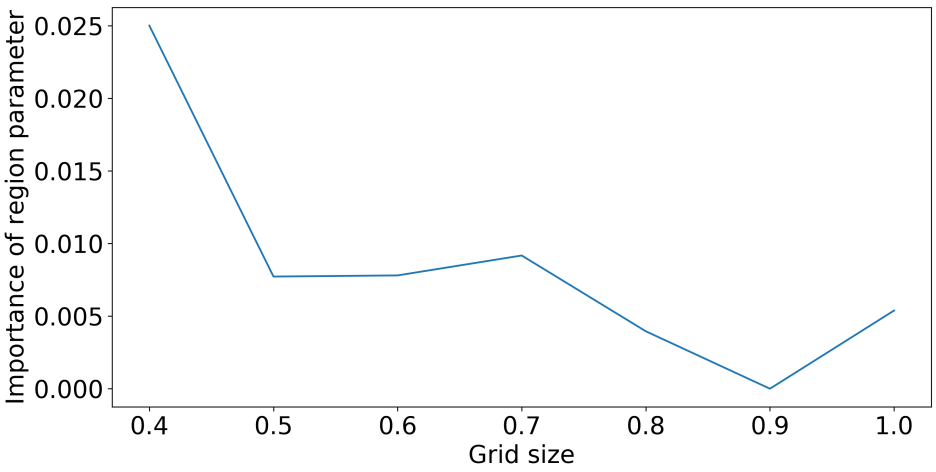

Figure S5.10: The importance of region parameter for different grid sizes.

Chapter 6

Conclusions and Future Work

6.1 Summary of Conclusions

In this dissertation, I have mapped the geometry of seismogenic faults and found that sedimentary faults and basement faults are under common tectonic control. I have carried out detailed seismogenic fault stress state analysis at both the state scale and local cluster scale. Evidences show that fault structure and fault stress state influence the rupture behaviors. I have explored the possible triggering mechanisms of earthquakes in Oklahoma, including pore pressure, poroelastic stress, and earthquake interactions. Although pore pressure and poroelastic stress are still the main triggering mechanisms of the induced earthquakes in Oklahoma, earthquake interactions are substantial within in clusters and can drive propagation of the sequences. Each of the factors not only has influence on the potential seismic hazard but also interact with each other, forming a more complex problem. For example, pore pressure increase from injection can affect the fault strength and cause changes in stress field through poroelastic effects.

Besides the four aspects (fault structure, fault stress state, earthquake interactions, injections) examined above, many other factors could also influence the rupture process and potential earthquake hazards. Rupture history could affect the seismic hazard assessment. Wu et al. (2018) found large spatial variability of stress drop in Oklahoma and did not support the idea of using low stress drops in ground-motion prediction models in central and eastern United States. Geological structures and rock properties such as permeability were found to affect the spatial location of the earthquakes (e.g. Shah & Keller, 2017; Pei et al., 2018). The studies (e.g., Kibikas et al., 2020; Kolawole et al., 2020) on physical properties of the basement rocks could help better understand and characterize the fault reactivation. Lab experiments also found aseismic slip through the coupling between fault slip and fluid flow (e.g., Cappa et al., 2019; Eyre et al., 2019). The detailed discussion of the above factors is beyond the research scope of this dissertation.

I have applied machine learning techniques to polarity picking and seismicity rate forecasting. The results have comparable precision as traditional methods, which demonstrates the potential of machine learning techniques in solving traditional problems in geophysics. More importantly, as shown in the last chapter, I started the seismicity forecasting as a statistical problem without any physical assumptions. As a result, the model could rank the input factors, and the ranking is consistent with our previous knowledge of the mechanisms behind the induced earthquakes. The results imply that the machine learning techniques could possibly find more hidden patterns in the data and the interpretations of the newly found patterns will promote our understanding of the physics for induced earthquakes.

6.2 Future Directions

Different methods have been used to assess the earthquake hazard in Oklahoma. The U.S. Geological Survey updated the earthquake hazard forecast in 2018 to account for the rapid changes in earthquake activity in the central and eastern U.S. (Petersen et al., 2018). The forecast worked well when compared to the observations (Brooks et al., 2018). The method to make the probabilistic forecast is mainly based on previous seismicity. With the known fault structures and fault stress state, more physical parameters can be used to assess potential earthquake hazard. Walsh III and Zoback (2016) utilized quantitative risk assessment (QRA) to calculate the conditional probability of slip on mapped faults in response to injection-related increases in pore pressure in northcentral Oklahoma by incorporating the uncertainty in each Mohr-Coulomb parameter (stress tensor, pore pressure, coefficient of friction, and fault orientation). Here in the first two chapters of this dissertation, I only qualitatively assessed the potential earthquake hazards based on the mapped fault geometries and stress state. In the following work, I will incorporate seismicity, fault geometry, stress state, and focal mechanisms in a machine learning model to make predictions of the maximum magnitude, earthquake number, and ground motions. The results will be compared to true observations to evaluate the model performance.

In the last chapter, the machine learning model generated decent forecast of the seismicity rate. The physics-based parameters had higher importance than the operational parameters. I also tried to incorporate a regional parameter in the input dataset, but without much success. To refine the model, I will further add some parameters to represent the spatial heterogeneity of seismicity. The

possible candidates include fault stress state, seismogenic index (Langenbruch et al., 2018), and a categorical parameter differentiating low- and high- velocities as in Pei et al. (2018). Hopefully the spatial parameter will better describe the various relationships between seismicity and injection in different areas, and the results will provide a better image of high-risk regions, which can serve as a reference to regulate future injection activities.

References

Brooks, E. M., Stein, S., Spencer, B. D., Salditch, L., Petersen, M. D., & McNamara, D. E. (2018). Assessing earthquake hazard map performance for natural and induced seismicity in the central and eastern united states. *Seismological Research Letters*, *89*(1), 118–126.

Cappa, F., Scuderi, M. M., Collettini, C., Guglielmi, Y., & Avouac, J.-P. (2019). Stabilization of fault slip by fluid injection in the laboratory and in situ. *Science advances*, *5*(3), eaau4065.

Eyre, T. S., Eaton, D. W., Garagash, D. I., Zecevic, M., Venieri, M., Weir, R., & Lawton, D. C. (2019). The role of aseismic slip in hydraulic fracturing–induced seismicity. *Science advances*, *5*(8), eaav7172.

Kibikas, W. M., Carpenter, B. M., & Ghassemi, A. (2020). Mechanical strength and physical properties of oklahoma's igneous basement. *Tectonophysics*, *777*, 228336.

Kolawole, F., Simpson Turko, M., & Carpenter, B. M. (2020). Basement-controlled deformation of sedimentary sequences, anadarko shelf, oklahoma. *Basin Research*.

Langenbruch, C., Weingarten, M., & Zoback, M. D. (2018). Physics-based forecasting of man-made earthquake hazards in oklahoma and kansas. *Nature communications*, *9*(1), 1–10. doi: 10.1038/s41467-018-06167-4

Pei, S., Peng, Z., & Chen, X. (2018). Locations of injection-induced earthquakes in oklahoma controlled by crustal structures. *Journal of Geophysical Research: Solid Earth*, *123*(3), 2332–2344.

Petersen, M. D., Mueller, C. S., Moschetti, M. P., Hoover, S. M., Rukstales, K. S., McNamara, D. E., ... others (2018). 2018 one-year seismic hazard forecast for the central and eastern united states from induced and natural earthquakes. *Seismological Research Letters*, *89*(3), 1049–1061.

Shah, A. K., & Keller, G. R. (2017). Geologic influence on induced seismicity: Constraints from potential field data in oklahoma. *Geophysical Research Letters*, *44*(1), 152–161.

Walsh III, F. R., & Zoback, M. D. (2016). Probabilistic assessment of potential fault slip related to injection-induced earthquakes: Application to north-central oklahoma, usa. *Geology*, *44*(12), 991–994.

Wu, Q., Chapman, M., & Chen, X. (2018). Stress-drop variations of induced earthquakes in oklahomastress-drop variations of induced earthquakes in oklahoma. *Bulletin of the Seismological Society of America*, *108*(3A), 1107–1123.