

Magnitude Conversion and Earthquake Recurrence Rate Models for the Central and Eastern United States

Date Published: March 2023

Prepared by:

Rasool Anooshehpour, U.S. Nuclear Regulatory Commission

Thomas Weaver, U.S. Nuclear Regulatory Commission

Jon Ake, U.S. Nuclear Regulatory Commission

Cliff Munson, U.S. Nuclear Regulatory Commission

Morgan Moschetti, U.S. Geological Survey

David Shelly, U.S. Geological Survey

Peter Powers, U.S. Geological Survey

Rasool Anooshehpour, NRC Program Manager

Research Information Letter
Office of Nuclear Regulatory Research

DISCLAIMER

Legally binding regulatory requirements are stated only in laws, NRC regulations, licenses, including technical specifications, or orders; not in Research Information Letters (RILs). A RIL is not regulatory guidance, although NRC's regulatory offices may consider the information in a RIL to determine whether any regulatory actions are warranted. This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, or any of their employees, makes any warranty, expressed or implied, or assumes any legal liability of responsibility for any third party's use, or the results of such use, or any information, apparatus, product or process disclosed in this report, or represents that its use by such third party would not infringe privately owned rights.

EXECUTIVE SUMMARY

Development of Seismic Source Characterization (SSC) models, which is an essential part of Probabilistic Seismic Hazard Analyses (PSHA), can help forecast the temporal and spatial distribution of future damaging earthquakes ($M_w \geq 5$) in seismically active regions. Because it is impossible to associate all earthquakes with known faults, seismic source models for PSHA often include sources of diffuse seismicity in which future earthquake scenarios are not localized on mapped faults. These sources of diffuse seismicity are referred to as area source zones, distributed seismicity zones, or just source zones.

In the Central and Eastern United States (CEUS) very few Quaternary-active faults have the requisite information for use in PSHA (i.e., fault geometry and dimensions, event rates or slip rates, etc.), and we lack knowledge about the causative faults for most observed seismicity in the region. As a result, area source zones are frequently used in site-specific PSHA in the CEUS to represent diffuse seismicity that cannot be associated with faults. However, there are examples of active fault sources in the CEUS, such as the Meers fault, the Cheraw fault, and New Madrid region, where individual faults can be characterized.

The source characterization models for background seismicity are based, to a large extent, on an assumption that spatial distribution of historical and recorded seismicity will not change substantially for time periods of interest for PSHA (approximately the next 50-100 years for engineered structures). Furthermore, correlation between the locations of small- to moderate-magnitude earthquakes and the locations of large-magnitude earthquakes, indicates that with some level of confidence, one can use the spatial pattern of smaller earthquakes to forecast the future pattern of damaging earthquakes.

Within background seismicity zones, the earthquake rate forecast is developed using spatial smoothing of the small to moderate magnitude events in earthquake catalogs. Different methodologies are used for this purpose and can predict varying distributions of seismicity rates. This in turn affects the results of a seismic hazard analysis. The U.S. Geological Survey (USGS) and Nuclear Regulatory Commission (NRC) use different methods for computing spatially smoothed seismicity rates in the CEUS; the USGS uses kernel-based spatial smoothing methods in developing the National Seismic Hazard Model (NSHM), and the method adopted in the Central and Eastern United States Seismic Source Characterization (CEUS-SSC) project is used when evaluating seismic hazard for nuclear power plant siting.

This RIL has been developed as part of RES activities to assess factors that affect estimates of background seismicity rates (i.e., earthquakes that are not associated with mapped faults) in the Central and Eastern United States (CEUS). This study was performed in part because previous NRC sponsored studies indicated that background seismicity rates estimated through different approaches by the USGS, and the US NRC may be the cause of the discrepancies between the two agencies seismic hazards estimates in the CEUS. In this collaborative research, background seismicity rates at three locations in the CEUS estimated by the USGS and NRC models were compared. Although the distribution of seismicity rates was generally similar, the study found that differences in the distribution may result in differences in mean hazard estimates.

In addition, background seismicity rates were found to be affected by models used to convert local magnitudes of small earthquakes to moment magnitudes. This study explored avenues to mitigate the current effects of magnitude inconsistencies through targeted studies to develop magnitude conversion relationships with lower bias, particularly focused on small magnitudes.

TABLE OF CONTENTS

DISCLAIMER	III
EXECUTIVE SUMMARY	V
TABLE OF CONTENTS	VII
LIST OF FIGURES	IX
LIST OF TABLES	XV
ABBREVIATIONS AND ACRONYMS	XVII
1 INTRODUCTION	1-1
2 EARTHQUAKE MAGNITUDES AND EARTHQUAKE CATALOG	2-1
2.1 Overview	2-1
2.2 Data and Results	2-2
2.2.1 Measurement Dataset.....	2-2
2.3 Magnitude Comparisons	2-3
2.4 Future Directions to Improve Magnitudes	2-8
3 BACKGROUND SEISMICITY SMOOTHING APPROACHES	3-1
3.1 Fixed and Adaptive Kernel Smoothing Approaches	3-1
3.2 Penalized-Likelihood Smoothing Approach	3-5
3.2.1 Likelihood Function	3-5
3.2.2 Penalty Function.....	3-10
3.3 Modeling the Joint Distribution of Recurrence Parameters	3-13
3.3.1 Development of Recurrence Maps	3-15
3.3.2 Alternative Recurrence Maps.....	3-17
3.3.3 Application of the Model and Specification of Model Parameters.....	3-22
3.3.4 Effect of Cell Size on Estimated Recurrence Rate.....	3-23
4 SMOOTHING APPROACH COMPARISONS	4-1
4.1 Catalog and Catalog Completeness	4-6
4.1.1 USGS Approach.....	4-7

4.1.2 CEUS-SSC Approach	4-8
4.2 Comparison of Recurrence Parameters to Catalog.....	4-11
4.2.1 Estimates of Earthquake Recurrence Parameters (Weichert, 1980).....	4-12
4.3 Hazard Curves	4-15
5 EVALUATIONS OF EPISTEMIC UNCERTAINTY	5-1
6 CONCLUSIONS	6-1
7 REFERENCES	7-1

LIST OF FIGURES

- Figure 2-1: Map of earthquakes in measurement dataset (dark blue dots), consisting of 399 events in Kansas and 361 events in Oklahoma. Labeled open black triangles indicate broadband seismic stations used in this study – stations beyond this map area also were used. Dashed line shows the Kansas/Oklahoma border. Gray box shows area from which events are selected (Figure from Shelly et al., 2021).2-3
- Figure 2-2: Magnitude comparison for M_L in northern Oklahoma. (A) Comparison for M_L magnitudes determined by the Oklahoma Geological Survey (OGS) (black “+”). Corresponding events with magnitudes converted using the Central and Eastern United States Seismic Source Characterization (CEUS-SSC) relationship are shown in blue. Red “x” indicates waveform-modeled M_W estimates from St. Louis University (SLU). (B) Catalog magnitude residuals from part (a) (after subtracting coda-envelope M_W). (C) Residuals for converted magnitudes from part (a). Symbols in parts (B) and (c) as in (a). Green diamonds in (B) and (C) show the median value of coda-envelope M_W for each M_L (or converted M_L) in increments of 0.1 magnitude units, which avoids biases from the $M_L \geq 2.5$ selection criterion (Figure from Shelly et al., 2021).2-4
- Figure 2-3: Magnitude comparison for M_L in southern Kansas. (A) Comparison for M_L magnitudes determined by the U.S. Geological Survey Induced Seismicity Menlo Park (ISMP) group. (black “+”). Corresponding events with magnitudes converted using the Central and Eastern United States Seismic Source Characterization (CEUS-SSC) relationship are shown in blue. Red “x” indicates waveform-modeled M_W estimates from St. Louis University (SLU). (B) Residuals for catalog magnitudes from part (A) (after subtracting coda-envelope M_W). (C) Residuals for converted magnitudes from part (A). Symbols in parts (B) and (C) as in (A). Green diamonds in (B) and (C) show the median value of coda-envelope M_W for each M_L (or converted M_L) in increments of 0.1 magnitude units, which avoids biases from the $M_L \geq 2.5 \geq$ selection criterion (Figure from Shelly et al., 2021).2-5
- Figure 2-4: Magnitude comparison for m_{bLg} in southern Kansas. (A) Comparison of ComCat m_{bLg} magnitudes determined by the US network (U.S. Geological Survey National Earthquake Information Center, black “+”) and associated magnitudes converted using the Central and Eastern United States Seismic Source Characterization (CEUS-SSC) m_{bLg} to M_W relationship (orange) or the Rigsby et al. (2014) relationship (purple). Red “x” indicates waveform-modeled M_W estimates from St. Louis University (SLU). (B) Magnitude residuals (after subtracting coda envelope M_W) for catalog magnitudes from (A). (C) Magnitude residuals (after subtracting coda envelope M_W) for converted magnitudes from (A). Symbols in parts (B) and (C) as in (A). Green diamonds in (B) and (C) show the median value of coda-envelope M_W for each m_{bLg} (or converted m_{bLg}) in increments of 0.1 magnitude units, which avoids biases due to the $m_{bLg} \geq 2.5$ selection criterion (Figure from Shelly et al., 2021).2-6

- Figure 2-5: Comparison of frequency-magnitude distributions and associated b -values using coda-derived M_W versus magnitudes converted from M_L to M_W using the Central and Eastern United States Seismic Source Characterization (CEUS-SSC) conversion relationship for this region (EPRI/DOE/NRC, 2012). Histograms show numbers of events for each magnitude bin. Blue line shows cumulative numbers of events at or above each magnitude. Dashed gray line shows the b -value slope, as determined by maximum likelihood (Utsu, 1966), with uncertainties calculated using the method of Shi and Bolt (1982). The completeness magnitude (M_c) used for b -value estimation is marked by the gray vertical line. (A) Northern Oklahoma dataset, using direct M_W estimates (waveform modeled if available, coda-derived for the remainder). (B) Same set of events as in (A) but using the converted (M_L to M_W) magnitudes for those lacking a waveform-modeled M_W . Note that this set of events is not fully complete above M_c in either case, but the difference in b -values between direct M_W and converted- M_W datasets is robust. (C) Same as (A), but for southern Kansas dataset. (D) same as (B), but for southern Kansas dataset. Figure from Shelly et al. (2021).....2-7
- Figure 3-1: Schematic representation of the location of earthquake j smoothed with the isotropic adaptive kernel, $K_{\sigma_j}(\vec{r}_j)$ is shown here. The density of seismicity in each cell is estimated by adding the contributions from all earthquake kernels in that cell.3-2
- Figure 3-2: Earthquake rate model for the Central and Eastern United States (CEUS) from an adaptive smoothed seismicity model. Annual rates are plotted in a log10 scale and represent the number of $M_W \geq 5.0$ earthquakes within each grid cell. The dashed line between western North America and Central and Eastern North America (CENA) delineates the western boundary of the CEUS for the purposes of this study (Figure 4 in Moschetti, 2015).....3-4
- Figure 3-3: Histogram of magnitudes in the earthquake catalog used in Central and Eastern United States Seismic Source Characterization (CEUS-SSC). The minimum magnitude shown (M_W 2.9) is the lowest magnitude used in these recurrence calculations (Figure 5.3.2-3, NUREG-2115, v2; NRC 2012). The red exponential curve highlights deviation of the data from a smooth exponential shape below M_W 3.5.....3-7
- Figure 3-4: A schematic representation of $f(x, y)$ in a cell, and in its four neighbors.3-11
- Figure 3-5: A schematic representation of $\beta(x, y)$ deviation from the local average.....3-12
- Figure 3-6: State vector, \mathbf{X} , of a source zone divided into M cells.3-13
- Figure 3-7: Plots of P -acceptance distribution as a function of $\ln v$, and v are shown here. Frame A shows normal distributions of $\ln v$ for mean values of 0.02, 0.002, and 0.0002. A standard deviation of 0.07 results in identical shapes for all three distributions. However, the same probabilities plotted as a function of recurrence rate v (frame B) have a different shape. These plots indicate that, with a fixed standard deviation for Markov Chain Monte Carlo (MCMC) trial distributions to generate a new state candidate, at regions with very low earthquake recurrence rates, the new realizations remain very close to the initial value. In contrast,

	where the recurrence rate is high, the new MCMC candidate could have a larger deviation from the current value.	3-15
Figure 3-8:	The North Appalachian (NAP), Non-Mesozoic-and-Younger Extension-Narrow (NMESE-N), and Paleozoic Extended Crust-Wide (PEZ-W) seismotectonic zones for which the recurrence rates were calculated. Note that PEZ-W is a sub-zone of NMESE-N.	3-16
Figure 3-9:	Maps of the mean recurrence rate ($M_W \geq 5.0$ events/year) and the b -values for the North Appalachian (NAP) seismotectonic zone estimated by using the Penalized-Likelihood smoothing method are shown on the left column. The grid size in this case is $0.1^\circ \times 0.1^\circ$	3-17
Figure 3-10:	Covariance matrix of a source zone divided into M cells. Due to symmetry, only the lower triangle (covariance) and the trace (variance) are shown here.	3-18
Figure 3-11:	A plot of the fraction of the cumulative variance (eigenvalues) that contribute to the total variance (trace of the covariance matrix) for the North Appalachian Seismictectonic Zone. There is a total of 788 cells ($0.25^\circ \times 0.25^\circ$). Here, the sum of the first 375 eigenvalues out of 1576 represents 99% of the total variance.	3-19
Figure 3-12:	An example of random generation of 6 sets of $\epsilon_j z$ is shown here. Each set (color-coded) is sampled from 8 different segments of the normal distribution of ϵ_j with equal probability of $1/8$	3-20
Figure 3-13:	The mean and eight alternative maps of earthquake recurrence rate ($M_W \geq 5.0$) for the North Appalachian seismotectonic zone are shown here. The eight alternative maps capture the central tendency and statistical uncertainty in the recurrence parameters. The grid size is $0.1^\circ \times 0.1^\circ$	3-21
Figure 3-14:	A comparison between the mean recurrence map ($M_W \geq 5$) and the average of the eight alternative maps for the NAP zone is shown here. The similarity between the two maps supports the notion that the alternative maps jointly represent the central tendency and statistical uncertainty in the recurrence parameters.	3-22
Figure 3-15:	Maps of the mean earthquake recurrence rates ($M_W \geq 5$ /yr) for Cases A, B, and E estimated by using the Penalized-Likelihood smoothing method for North Appalachian (NAP) zone. The grid size in this case is $0.1^\circ \times 0.1^\circ$	3-23
Figure 3-16:	Mean recurrence rate maps of North Appalachian (NAP) zone calculated using three different grid sizes for (A) Penalized-Likelihood and (B) Kernel approaches. In general, the rate distributions remain the same with higher resolution at smaller grid size. However, in Case E (Panel A), the rate of smoothing tends to increase (more uniform) as the grid size increases.	3-24
Figure 4-1:	This figure shows differences in estimates of the recurrence rates for $M_W \geq 5.0$ earthquakes in North Appalachian (NAP) using the adaptive kernel (A), the fixed kernel (B-D), and the penalized-likelihood (E-G) smoothing approaches. The chart shows the total zone rates ($M_W \geq 5.0$) for all approaches. The error bars show the range covered by the eight	

	realizations in the penalized-likelihood method. In all cases, the 2018 U.S. Geological Survey (USGS) earthquake catalog is used as the input.	4-2
Figure 4-2:	This figure shows differences in estimates of the recurrence rates for $M_W \geq 5.0$ earthquakes in Paleozoic Extended Crust-Wide (PEZ-W) using the same smoothing approaches and inputs for each subpanel A-G as described for Figure 4-1.	4-2
Figure 4-3:	This figure shows differences in estimates of the recurrence rates for $M_W \geq 5.0$ earthquakes in Non-Mesozoic-and-Younger Extension-Narrow (NMESE-N) using the same smoothing approaches and inputs for each subpanel A-G as described for Figure 4-1.	4-3
Figure 4-4:	This figure depicts estimates of the recurrence rates for $M_W \geq 5.0$ earthquakes in North Appalachian (NAP) between Nuclear Regulatory Commission (NRC) and U.S. Geological Survey (USGS) rate models. NRC rate models include Case A, Case B, and Case E. USGS rate models include the adaptive smoothing (ad) and fixed smoothing with $M_c = 2.7$ (fixR-M2.7), $M_c = 3.7$ (fixR-M3.7), and $M_c = 4.7$ (fixR-M4.7).	4-4
Figure 4-5:	This figure depicts estimates of the recurrence rates for $M_W \geq 5.0$ earthquakes in Paleozoic Extended Crust-Wide (PEZ-W) between Nuclear Regulatory Commission (NRC) and U.S. Geological Survey (USGS) rate models. All other model details are the same as in Figure 4-4.	4-5
Figure 4-6:	This figure depicts estimates of the recurrence rates for $M_W \geq 5.0$ earthquakes in Non-Mesozoic-and-Younger Extension-Narrow (NMESE-N) between Nuclear Regulatory Commission (NRC) and U.S. Geological Survey (USGS) rate models. All other model details are the same as in Figure 4-4.	4-6
Figure 4-7:	The seven zones of completeness (ZC_1 – ZC_6 , ZC_{out}) used in the U.S. Geological Survey (USGS) smoothing approach to generate earthquake recurrence rate maps. The tables list M_c and the years of catalog completeness for both fixed and adaptive kernel methods in each zone.	4-8
Figure 4-8:	A schematic showing how catalog completeness is determined through Stepp (1972) approach. Dates for t_p and t_b are based on the 2018 U.S. Geological Survey (USGS) catalog.	4-8
Figure 4-9:	The “Stepp” plot of earthquake recurrence rate is shown in (A) as a function of time for catalog completeness in Region 12 (Figure 3.5-4, NUREG-2115; NRC 2012). The 14 regions of completeness designated for the Central and Eastern United States Seismic Source Characterization (CEUS-SSC) are shown in (B). The table shows the “beginning year of usable data” and the updated “equivalent time of completeness” in Region 12 using the U.S. Geological Survey (USGS) 2018 earthquake catalog (Case B, Table 3-2).	4-10
Figure 4-10:	Maps in the left column show the mean earthquake recurrence rates ($M_W \geq 5$) in the North Appalachian (NAP) seismotectonic region calculated using the Penalized-Likelihood approach for Case A (see Table 3-2). Recurrence rates in (A-C) are calculated using the seismic	

zones and completeness periods of the Central and Eastern United States Seismic Source Characterization (CEUS-SSC), U.S. Geological Survey (USGS) Adaptive Kernel (AK), and USGS Fixed Kernel (FK), respectively. Although there are some differences, the distributed seismicity looks very similar in all three maps. This is supported by the chart on the right, which shows almost identical total recurrence rates in NAP for all three cases.....4-11

- Figure 4-11: Likelihood function for b -value of an exponential magnitude distribution, for multiple values of the earthquake count N . The value of b is normalized by the maximum-likelihood estimate b_{ML} . (Figure 5.3.2-2, NUREG-2115; NRC 2012).....4-12
- Figure 4-12: Penalized-likelihood predicted b -values for the eight recurrence rate realizations in the North Appalachian (NAP) and Paleozoic Extended Crust-Wide (PEZ-W) zones using Cases A, B, and E magnitude weights are compared with those obtained from the Weichert (1980) approach and using the equivalent period of completeness, T_E , in place of Weichert's unequal observational period, t_c^j in Equation 4-4. The error bars represent the 16%–84% uncertainty associated with the earthquake data (Weichert, 1980). For comparison, the unity b -values the USGS adaptive- and fixed- kernel approaches also are shown. (Note that the 'Weichert b ' line is drawn to show the slope only, not the actual magnitude distribution.)4-14
- Figure 4-13: Hazard curves computed at 10 Hz spectral acceleration at a select site in the North Appalachian (NAP) seismotectonic zone (see Table 4-2 for coordinates) using Adaptive Kernel (AK), Fixed Kernel (FK), and Penalized-Likelihood (PL) methods. Maps show results computed with different approaches.....4-16
- Figure 4-14: Hazard curves computed at 10 Hz spectral acceleration at a select site in the Paleozoic Extended Crust-Wide (PEZ-W) seismotectonic zone (see Table 4-2 for coordinates) using Adaptive Kernel (AK), Fixed Kernel (FK), and Penalized-Likelihood (PL) methods. Maps show results computed with different approaches.....4-16
- Figure 4-15: Hazard curves computed at 10 Hz spectral acceleration at a select site in Non-Mesozoic-and-Younger Extension-Narrow (NMESE-N) seismotectonic zone (see Table 4-2 for coordinates) using Adaptive Kernel (AK), Fixed Kernel (FK), and Penalized-Likelihood (PL) methods. Maps show results computed with different approaches.4-17
- Figure 4-16: Various combinations of the mean and weighted mean hazard curves computed at 10 Hz spectral acceleration at select sites in North Appalachian (NAP), Paleozoic Extended Crust-Wide (PEZ-W), and Non-Mesozoic-and-Younger Extension-Narrow (NMESE-N) seismotectonic zones are plotted here. The frames on the left show the mean hazard curves for Penalized-Likelihood (PL), Adaptive Kernel (AK), and three Fixed Kernel (FK) methods. The middle frames show PL mean (average of 24 rate realizations), AK mean, and the weighted mean of FK (average of three different FK approaches). The frames on the right show the PL mean and the kernel weighted mean ($0.4 \times AK + 0.6 \times FK$) hazards.....4-18

Figure 5-1: Total seismicity rates for North Appalachian (NAP), Paleozoic Extended Crust-Wide (PEZ-W), and Non-Mesozoic-and-Younger Extension-Narrow (NMESE-N) seismotectonic zones. The error bars indicate the range of total seismicity rates for eight realizations in each of the three cases (A, B, and E).5-3

Figure 5-2: North Appalachian (NAP) 10 Hz hazard curves comparing penalized-likelihood (PL) results with results using (A) adaptive kernel (AK), (B) fixed kernel (FK) with a minimum magnitude of 2.7, (C) fixed kernel with a minimum magnitude of 3.7, and (D) fixed kernel with a minimum magnitude of 4.7.....5-4

Figure 5-3: Paleozoic Extended Crust-Wide (PEZ-W) 10 Hz hazard curves comparing penalized-likelihood (PL) results with results using (A) adaptive kernel (AK), (B) fixed kernel (FK) with a minimum magnitude of 2.7, (C) fixed kernel with a minimum magnitude of 3.7, and (D) fixed kernel with a minimum magnitude of 4.7.5-5

Figure 5-4: Non-Mesozoic-and-Younger Extension-Narrow (NMESE-N) 10 Hz hazard curves comparing penalized-likelihood (PL) results with results using (A) adaptive kernel (AK), (B) fixed kernel (FK) with a minimum magnitude of 2.7, (C) fixed kernel with a minimum magnitude of 3.7, and (D) fixed kernel with a minimum magnitude of 4.7.5-6

Figure 5-5: The mean and weighted mean hazard curves at select sites in North Appalachian (NAP), Paleozoic Extended Crust-Wide (PEZ-W), and Non-Mesozoic-and-Younger Extension-Narrow (NMESE-N) seismotectonic zones are plotted here for Penalized-Likelihood (PL), Adaptive Kernel (AK), and Fixed Kernel (FK) methods.....5-7

LIST OF TABLES

Table 3-1:	An example of input (training) and target (testing) catalog information used for evaluating model event density μ and log-likelihood LL , respectively (from Helmstetter et al., 2007).	3-3
Table 3-2:	Alternative cases considered for the magnitude-dependent weights.	3-22
Table 4-1:	Probability of detection and equivalent periods of completeness for Region 12 (Figure 4-9) for Case B (extracted from Table 3.5-2, NUREG 2115; NRC 2012).	4-9
Table 4-2:	Locations of sites where hazard curves were calculated.....	4-15

ABBREVIATIONS AND ACRONYMS

ADAMS	Agency-wide Documents Access and Management System
AK	Adaptive Kernel
CCT	Coda Calibration Tool
CENA	Central and Eastern North America
CEUS	Central and Eastern United States
CEUS-SSC	Central and Eastern United States Seismic Source Characterization
ComCat	Comprehensive Earthquake Catalog
DOE	Department of Energy
EPRI	Electric Power Research Institute
FK	Fixed Kernel
GMM	Ground Motion Model
GMRS	Ground Motion Response Spectrum
ISMP	Induced Seismicity Menlo Park
MCMC	Markov Chain Monte Carlo
NAP	North Appalachian
NEIC	National Earthquake Information Center
NGA-East	Next Generation Attenuation Relationships for Central & Eastern North America
NMESE-N	Non-Mesozoic-and-Younger Extension-Narrow
NRC	Nuclear Regulatory Commission
NSHM	National Seismic Hazard Model
NSHMP	National Seismic Hazard Modeling Project
NUREG	Nuclear Regulatory Report
OGS	Oklahoma Geological Survey
PL	Penalized-Likelihood
PEZ-W	Paleozoic Extended Crust-Wide
PSHA	Probabilistic Seismic Hazard Analyses
RIL	Research Information Letter
RLME	Repeated Large Magnitude Earthquake
SLU	Saint Louis University
SOG	Seismicity Owners Group
SSC	Seismic Source Characterization
SSHAC	Senior Seismic Hazard Analysis Committee
UCERF3	Uniform California Earthquake Rupture Forecast, Version 3
USGS	United States Geological Survey
TE	Equivalent Period of Completeness

1 INTRODUCTION

Development of Seismic Source Characterization (SSC) models, which is an essential part of Probabilistic Seismic Hazard Analyses (PSHA), can help forecast the temporal and spatial distribution of future damaging earthquakes ($M_w \geq 5$) in seismically active regions. Because it is impossible to associate all earthquakes with known faults, seismic source models for PSHA often include sources of diffuse seismicity in which future earthquake scenarios are not localized on mapped faults. These sources of diffuse seismicity are referred to as area source zones, distributed seismicity zones, or just source zones. During the early years of PSHA studies, it was assumed that earthquakes in seismotectonic zones have (1) uniform spatial distribution, (2) Poisson temporal distribution, and (3) exponential magnitude distribution (NRC, 2012). In seismically active regions (e.g., the Western United States), where active faults are readily identified, models of the spatial distribution of earthquakes include both the fault source geometries and the distributed seismicity (background) source zones. Source characterization of active faults is complemented by paleoseismic studies with estimates of earthquake magnitudes, dates of occurrences, and slip rates, which provide important information for PSHA studies.

In the Central and Eastern United States (CEUS) very few Quaternary-active faults have the requisite information for use in PSHA (i.e., fault geometry and dimensions, event rates or slip rates, etc.), and we lack knowledge about the causative faults for most observed seismicity in the region. As a result, area source zones are frequently used in site-specific PSHA in the CEUS to represent diffuse seismicity that cannot be associated with faults. However, there are examples of active fault sources in the CEUS, such as the Meers fault, the Cheraw fault, and New Madrid region, where individual faults can be characterized.

The source characterization models for background seismicity are based, to a large extent, on an assumption that spatial distribution of historical and recorded seismicity will not change substantially for time periods of interest for PSHA (approximately the next 50-100 years for engineered structures). Furthermore, studies such as those by Kafka (2007, 2009) found a correlation between the locations of small- to moderate-magnitude earthquakes and the locations of large-magnitude earthquakes, indicating that we can, with some level of confidence, use the spatial pattern of smaller earthquakes to forecast the future pattern of damaging earthquakes.

Within background seismicity zones, the earthquake rate forecast is developed using spatial smoothing of the small to moderate magnitude events in earthquake catalogs. Different methodologies are used for this purpose and can predict varying distributions of seismicity rates. This in turn affects the results of a seismic hazard analysis. The U.S. Geological Survey (USGS) and Nuclear Regulatory Commission (NRC) use different methods for computing spatially smoothed seismicity rates in the CEUS; the USGS uses kernel-based spatial smoothing methods in developing the National Seismic Hazard Model (NSHM), and the method adopted in the Central and Eastern United States Seismic Source Characterization (CEUS-SSC) project is used when evaluating seismic hazard for nuclear power plant siting. These methods are described and the impact on seismic hazard are evaluated in this Research Information Letter (RIL).

Another important input to estimating the rate of distributed seismicity is event magnitudes listed in earthquake catalogs. A substantial source of uncertainty in catalogs is the magnitude assigned to a given earthquake. Numerous different magnitude types exist, with each magnitude type computed in a different way. Therefore, for the sake of consistency, both the CEUS-SSC and the USGS NSHM have attempted to assemble a complete catalog with a uniform magnitude determination. To this end, moment magnitude, M_w , which is a physics-based measurement, has

been adopted as the standard. However, M_w was not computed routinely until the past few decades. To address this issue, the CEUS-SSC conducted extensive analyses to determine conversion equations from which to take a routinely computed network (e.g., M_L or m_{bLg}) and convert it into M_w . Another issue with using M_w is that it becomes increasingly difficult to compute for earthquakes with M_w less than ~ 4 .

This study investigates the effects of moment magnitude estimation and spatial smoothing methods on estimation of the earthquake rate forecast and on seismic hazard. We investigate the validity of the magnitude conversion equations and their associated uncertainties by applying them to a case study for induced earthquakes in southern Kansas and northern Oklahoma, and summarize the use of the decay of the seismic coda to estimate M_w for small earthquakes ($M_w < 4$). Furthermore, the study documents a comparison and assessment of background seismicity smoothing methods implemented by the USGS for the NSHM and used by the CEUS-SSC for siting nuclear facilities based on probabilistic seismic hazard estimates from multiple source zones in the CEUS and for multiple sites.

2 EARTHQUAKE MAGNITUDES AND EARTHQUAKE CATALOG

2.1 Overview

Seismic hazard forecasts are central to the Nuclear Regulatory Commission (NRC)'s mission of nuclear safety. One important input to seismic hazard forecasts is the historical rate of occurrence of earthquakes of certain magnitude in a particular area. This is based on catalogs of earthquake locations and magnitudes occurring over time. However, a substantial source of uncertainty in such determinations is the magnitude assigned to a given earthquake.

Earthquake magnitudes are intended to be measures of earthquake size. However, numerous different magnitude types exist, with each magnitude type computed in a different way. Although the numerical magnitudes from different magnitude types are often similar, they can differ in important ways. Even magnitudes that are nominally the same type (e.g., local magnitude, M_L) can differ depending on the details of the data selection and computation procedure, which often vary among different monitoring networks. Furthermore, seismic instrumentation, preferred magnitude type, and computation methods all may change over time. In addition to potentially introducing biases, these issues also complicate attempts to determine a “magnitude of completeness” above which the catalog is complete, which is usually a requirement for computing the associated Gutenberg-Richter b -value (Section 4.1). The b -value, which represents the slope of the frequency magnitude distribution, is an important parameter used to extrapolate the observed rate of small earthquakes to estimate the long-term rate of large earthquakes.

These variations in earthquake magnitude make it difficult to compare earthquake magnitudes or rates in different areas or even for different times. This problem has been recognized previously and attempts to address it have been made. For example, both the CEUS-SSC and the USGS NSHM go through substantial effort to attempt to assemble a complete catalog with a uniform magnitude determination. Toward this goal, moment magnitude, M_w , has been adopted as the standard. Although most magnitude types are defined by a measurement procedure, moment magnitude has the advantage that it is based on a physical quantity of the earthquake source, that is, the seismic moment. Moment magnitude has become the preferred magnitude with which to characterize large and moderate earthquakes ($M_w > \sim 4$), yet moment magnitude becomes increasingly difficult to compute for smaller earthquakes and is not routinely computed for earthquakes with magnitude less than ~ 4 . Furthermore, M_w was not computed routinely until the past few decades. So, it remains a challenge to obtain reliable estimates of M_w for small and/or historical earthquakes.

To address this issue, the CEUS-SSC conducted extensive analysis to determine conversion equations from which to take a routinely computed network magnitude (e.g., M_L or m_{bLg}) and convert it into M_w , aimed at being valid for the entire CEUS (EPRI/DOE/NRC, 2012). Although these equations take into account differences based on region, and in a few cases based on monitoring networks and time, they cannot account fully for temporal or spatial variations. Furthermore, the available M_w magnitudes from which these relations were derived were mostly limited to larger earthquakes ($M_w > 4$). Therefore, they may not perform well for smaller earthquakes.

Motivated by the potential effect on seismic hazard assessments, we performed a case study for earthquakes in southern Kansas and northern Oklahoma. Although this is an area of primarily induced seismicity, we expect the magnitude-related issues encountered in this area to be qualitatively similar to areas of purely tectonic seismicity. To determine M_w for earthquakes <

M_w 4, we applied a technique that utilizes the decay of the seismic coda, consisting of scattered seismic waves, which is well-suited to determining M_w for small earthquakes (Mayeda et al., 2003). However, before it can be reliably applied, this technique requires calibration for seismic stations and earthquakes in the region of interest. We performed such a calibration to determine M_w for 399 earthquakes as low as $M_w \sim 2$ in southern Kansas and 361 in northern Oklahoma. We then compared these moment magnitudes with the routine network magnitudes. Furthermore, we compared the moment magnitudes with those that would have been obtained based on the CEUS-SSC conversion relations. The results are presented in Shelly et al. (2021) and in the sections that follow.

As expected, we found significant differences between routinely computed M_L and m_{bLg} magnitudes. We also found that the CEUS-SSC conversion relationships perform poorly for this dataset, especially those events with $M_w < 3.5$. In fact, application of these conversion relationships, as published, often resulted in increased rather than decreased bias, compared to the raw routine magnitudes. Although some scatter was expected, we observe large systematic biases in magnitude for the converted versus computed M_w , in some cases up to 0.5 magnitude units. This is important because a systematic bias of 0.5 magnitude units corresponds to approximately a factor of three error in seismicity rates. Furthermore, b -values, which describe the relative fraction of large versus small earthquakes, are also affected by biases introduced in the conversion relationship. In the seismic hazard estimation using the CEUS-SSC model, b -values are constrained near 1.0, so the potential seismicity rate bias may be the larger influence on current seismic hazard analyses.

The following text and figures (Sections 2.2-2.5) summarize the results originally published in Shelly et al. (2021). Details of the applied techniques and further discussion can be found in Shelly et al. (2021).

2.2 Data and Results

By examining both southern Kansas and northern Oklahoma, we were able to compare routine catalog magnitudes from geologically similar areas calculated with slightly different procedures. Local magnitudes were routinely calculated for Kansas events by the USGS in Menlo Park, California (Rubinstein et al., 2018). Local magnitudes for Oklahoma events were calculated by the Oklahoma Geological Survey (OGS) (Darold et al., 2014; Walter et al., 2020). Together, southern Kansas and northern Oklahoma provide a dataset for comparing magnitudes of small earthquakes.

2.2.1 Measurement Dataset

After completing the calibration, we then applied the calibration to a much larger dataset, consisting of smaller-magnitude events. For this dataset, we selected events with USGS Comprehensive Earthquake Catalog (ComCat) (<https://earthquake.usgs.gov/earthquakes/search/>, last accessed 8 November 2020) preferred magnitude (in this case, M_L , m_{bLg} , or M_w) ≥ 2.5 . For southern Kansas, we selected events from the region shown in Figure 2-1 occurring from July 2014 to June 2019, when the bulk of the stations we used were active. Because of a higher rate of seismicity in Oklahoma, we limited these events to between November 2015 and June 2019. We followed the same initial processing steps that were applied to the smaller calibration dataset (waveform download, instrument response removal, envelope creation, and averaging on horizontal components). We then used the Coda Calibration Tool (CCT; Barno, 2017) in measurement-only mode to measure

M_w magnitudes for this dataset, by applying the calibration developed from the subset of larger magnitude events described above. We successfully determined a coda-based M_w for 399/405 events attempted in Kansas and 361/369 events attempted in Oklahoma, with almost all events falling between M_w 2.0 and M_w 4.0.

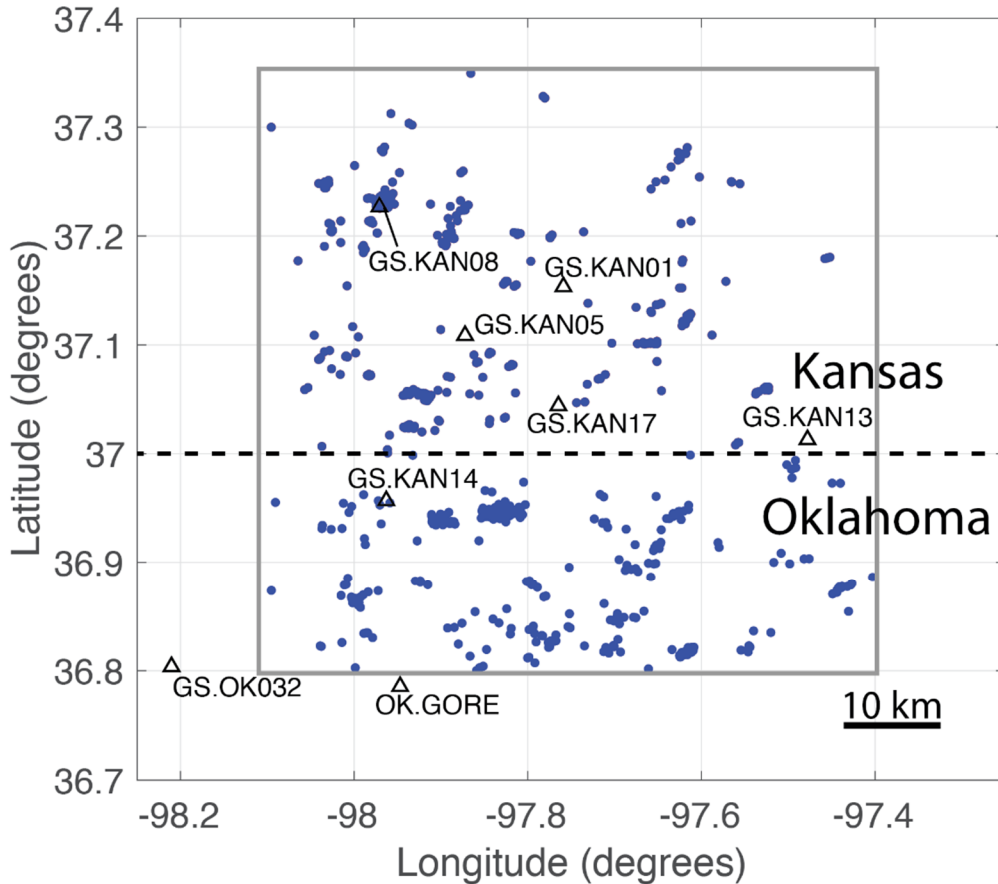


Figure 2-1: Map of earthquakes in measurement dataset (dark blue dots), consisting of 399 events in Kansas and 361 events in Oklahoma. Labeled open black triangles indicate broadband seismic stations used in this study – stations beyond this map area also were used. Dashed line shows the Kansas/Oklahoma border. Gray box shows area from which events are selected (Figure from Shelly et al., 2021).

2.3 Magnitude Comparisons

Comparisons of the coda-envelope-derived magnitudes with other magnitudes are shown in Figure 2-2 to Figure 2-4. We first note that waveform-modeled M_w magnitudes from Saint Louis University (SLU) (Herrmann et al., 2011) are generally consistent with coda-derived M_w , as expected from our calibration procedure. Next, we examine the relationship between commonly reported catalog magnitudes and coda-derived M_w magnitudes. Figure 2-3

For Oklahoma, we compared the coda-derived M_w magnitudes with catalog M_L reported by the OGS (Darold et al., 2014; Walter et al., 2020). As shown in Figure 2-2, we find a systematic difference between M_L and M_w of $\sim 0.3 - 0.4$ magnitude units (with large scatter), even at $M \sim 3$

where M_L and M_w are sometimes assumed to be similar. We note that OGS changed procedures used to calculate M_L in late-2018 (Walter et al., 2020); all but three magnitudes utilized here were calculated under the earlier procedure (Darold et al., 2014). Because of this change, Oklahoma event magnitudes may need to be treated differently for the earlier (pre-late-2018) versus later (post-late-2018) datasets. However, with only three post-change events included in our dataset (<1% of the total), we lack the dataset to constrain the later time period and simply analyze the events together here. OGS M_L magnitudes were reported to the nearest tenth of a magnitude unit before this procedure change, and to the nearest hundredth after the change; we analyze the magnitudes as reported.

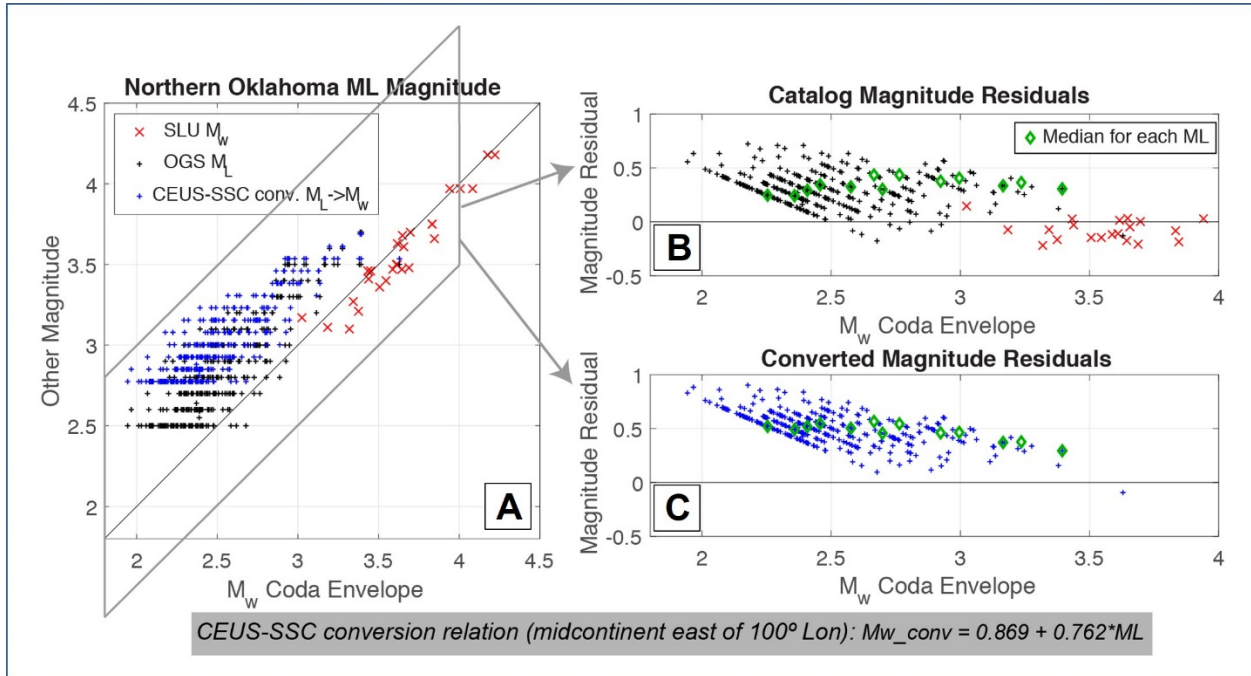


Figure 2-2: Magnitude comparison for M_L in northern Oklahoma. (A) Comparison for M_L magnitudes determined by the Oklahoma Geological Survey (OGS) (black “+”). Corresponding events with magnitudes converted using the Central and Eastern United States Seismic Source Characterization (CEUS-SSC) relationship are shown in blue. Red “x” indicates waveform-modeled M_w estimates from St. Louis University (SLU). (B) Catalog magnitude residuals from part (a) (after subtracting coda-envelope M_w). (C) Residuals for converted magnitudes from part (a). Symbols in parts (B) and (c) as in (a). Green diamonds in (B) and (C) show the median value of coda-envelope M_w for each M_L (or converted M_L) in increments of 0.1 magnitude units, which avoids biases from the $M_L \geq 2.5$ selection criterion (Figure from Shelly et al., 2021).

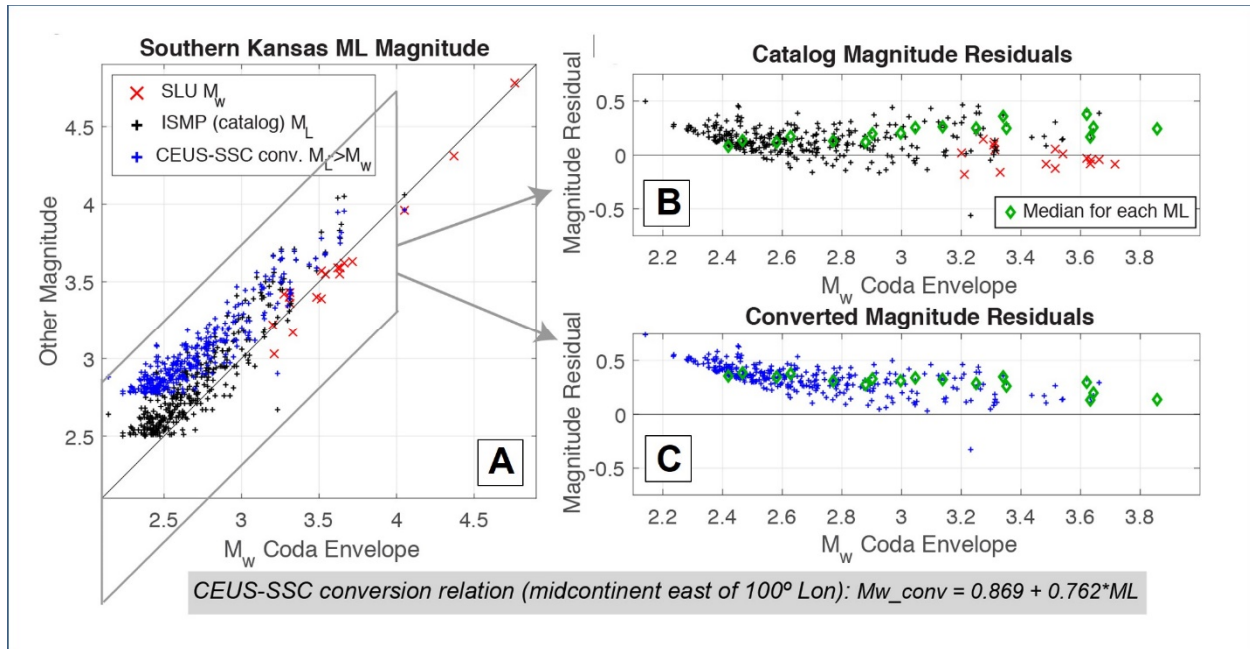


Figure 2-3: Magnitude comparison for M_L in southern Kansas. (A) Comparison for M_L magnitudes determined by the U.S. Geological Survey Induced Seismicity Menlo Park (ISMP) group. (black “+”). Corresponding events with magnitudes converted using the Central and Eastern United States Seismic Source Characterization (CEUS-SSC) relationship are shown in blue. Red “x” indicates waveform-modeled M_w estimates from St. Louis University (SLU). (B) Residuals for catalog magnitudes from part (A) (after subtracting coda-envelope M_w). (C) Residuals for converted magnitudes from part (A). Symbols in parts (B) and (C) as in (A). Green diamonds in (B) and (C) show the median value of coda-envelope M_w for each M_L (or converted M_L) in increments of 0.1 magnitude units, which avoids biases from the $M_L \geq 2.5$ selection criterion (Figure from Shelly et al., 2021).

Routine M_L magnitudes were calculated for the Kansas events by the USGS Induced Seismicity project in Menlo Park (ISMP) (Rubinstein et al., 2018). As shown in Figure 2-4, M_L magnitudes average ~ 0.1 magnitude units higher than the coda-envelope-derived M_w magnitudes at $M_L \sim 2.5$, increasing to ~ 0.3 magnitude units at $M_L \sim 3.5$, although scatter is considerable. This result is qualitatively similar to Oklahoma, but with slightly smaller offset. The difference from Oklahoma M_L is likely because of different procedures used by OGS compared to the ISMP M_L in Kansas, including the use of a different attenuation relationship (Darold et al., 2014; Greig et al., 2018; Al-Ismaïl et al., 2020).

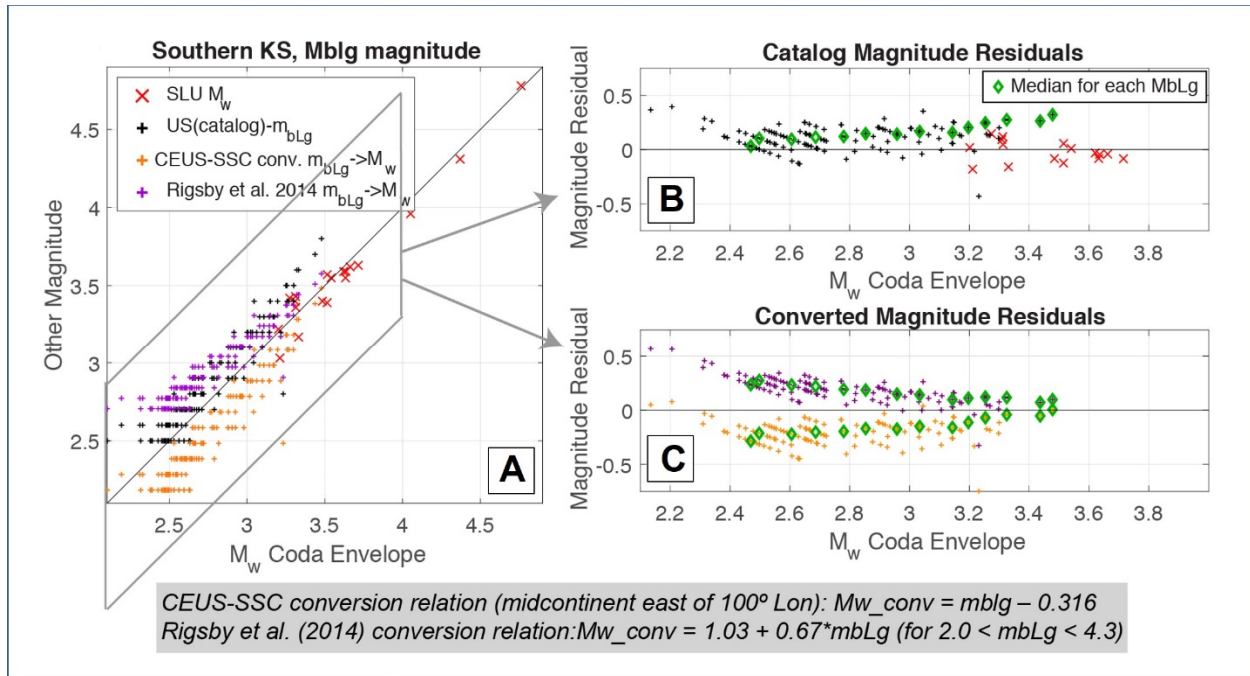


Figure 2-4: Magnitude comparison for m_{bLg} in southern Kansas. (A) Comparison of ComCat m_{bLg} magnitudes determined by the US network (U.S. Geological Survey National Earthquake Information Center, black “+”) and associated magnitudes converted using the Central and Eastern United States Seismic Source Characterization (CEUS-SSC) m_{bLg} to M_w relationship (orange) or the Rigsby et al. (2014) relationship (purple). Red “x” indicates waveform-modeled M_w estimates from St. Louis University (SLU). (B) Magnitude residuals (after subtracting coda envelope M_w) for catalog magnitudes from (A). (C) Magnitude residuals (after subtracting coda envelope M_w) for converted magnitudes from (A). Symbols in parts (B) and (C) as in (A). Green diamonds in (B) and (C) show the median value of coda-envelope M_w for each m_{bLg} (or converted m_{bLg}) in increments of 0.1 magnitude units, which avoids biases due to the $m_{bLg} \geq 2.5$ selection criterion (Figure from Shelly et al., 2021).

The last catalog magnitude we examined was the m_{bLg} magnitude for a subset of Kansas events, as computed by the USGS National Earthquake Information Center (NEIC) (see Rigsby et al., 2014 for magnitude computation details). The m_{bLg} magnitude is an adaptation of the magnitude developed by Nuttli (1973, 1986) based on the propagation of the Lg phase in central and eastern United States. m_{bLg} is commonly the preferred magnitude for events in the CEUS, when directly computed M_w is not available. As shown in Figure 2-4, these magnitudes track coda M_w well, with relatively small scatter, but they are shifted systematically higher by an average of ~ 0.1 magnitude units. Figure 2-2 to Figure 2-4 also show the results after applying magnitude conversion relationships that have been developed for the CEUS.

Finally, Figure 2-5 shows the differences in Gutenberg-Richter b -values between the direct M_w magnitudes and the M_w magnitude derived from applying the CEUS-SSC conversion relationships to the routine M_L magnitudes.

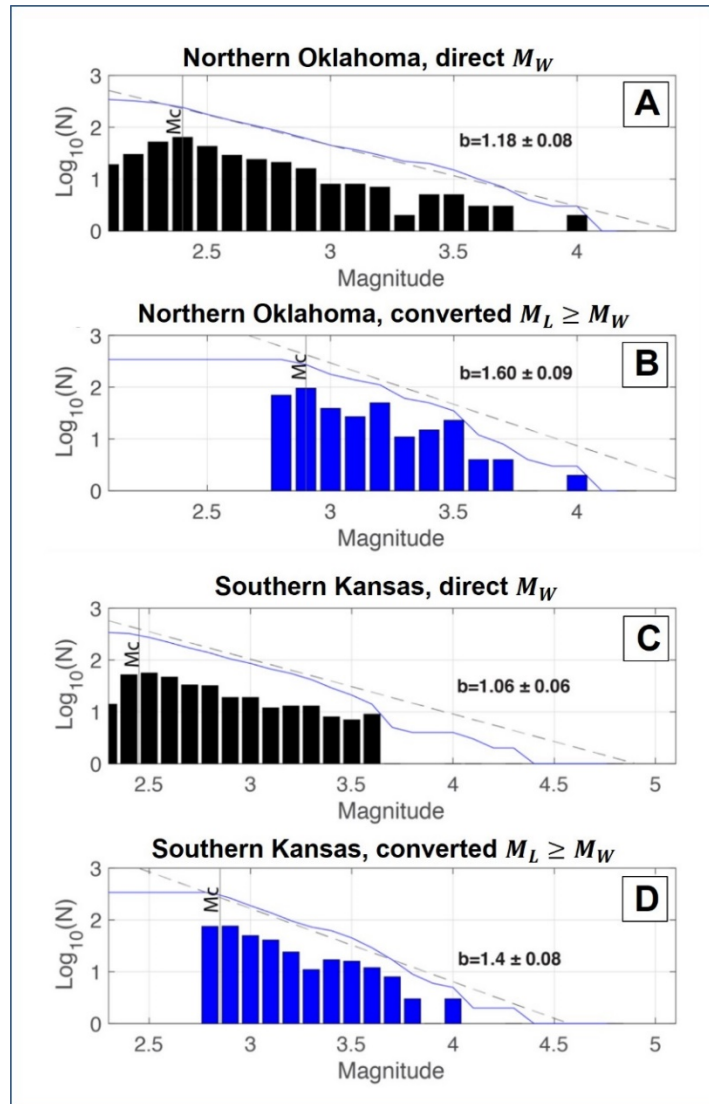


Figure 2-5: Comparison of frequency-magnitude distributions and associated b -values using coda-derived M_w versus magnitudes converted from M_L to M_w using the Central and Eastern United States Seismic Source Characterization (CEUS-SSC) conversion relationship for this region (EPRI/DOE/NRC, 2012). Histograms show numbers of events for each magnitude bin. Blue line shows cumulative numbers of events at or above each magnitude. Dashed gray line shows the b -value slope, as determined by maximum likelihood (Utsu, 1966), with uncertainties calculated using the method of Shi and Bolt (1982). The completeness magnitude (M_c) used for b -value estimation is marked by the gray vertical line. (A) Northern Oklahoma dataset, using direct M_w estimates (waveform modeled if available, coda-derived for the remainder). (B) Same set of events as in (A) but using the converted (M_L to M_w) magnitudes for those lacking a waveform-modeled M_w . Note that this set of events is not fully complete above M_c in either case, but the difference in b -values between direct M_w and converted- M_w datasets is robust. (C) Same as (A), but for southern Kansas dataset. (D) same as (B), but for southern Kansas dataset. Figure from Shelly et al. (2021).

2.4 Future Directions to Improve Magnitudes

Shelly et al. (2021) proposed three avenues for future mitigation of magnitude inconsistencies, including (1) greater acknowledgement of these uncertainties, (2) targeted studies to develop magnitude conversion relationships with lower bias for small earthquakes, valid for specific regions, monitoring networks, and time periods, and (3) further exploration of methods for routine computation of M_w for small earthquakes. Future exploration of the sensitivities of the seismic hazard models to these magnitude inconsistencies would also help target work where the benefit was greatest.

3 BACKGROUND SEISMICITY SMOOTHING APPROACHES

The CEUS-SSC model is based to a large extent on an assumption that spatial stationarity of seismicity will persist for time periods of interest for PSHA (approximately the next 50-100 years for engineered structures). Stationarity in this sense does not mean that future locations and magnitudes of earthquakes will occur exactly where they have occurred in the historical and instrumental record. Rather, the degree of spatial stationarity varies as a function of the type of data available to define the seismic source.

Patterns of seismicity away from known faults within seismically active zones (background seismicity) are defined from generally small- to moderate-magnitude earthquakes that have occurred during a relatively short (i.e., relative to the repeat times of large events) time period and cataloged as historical and instrumental records. Thus, where the locations of future events are not as tightly constrained by the locations of known fault sources, the earthquake rate forecast is developed using spatial smoothing of earthquake locations identified in earthquake catalogs. Different methodologies are used for this purpose. Below, we describe commonly used approaches in the CEUS to determine earthquake rates within a single seismic source.

3.1 Fixed and Adaptive Kernel Smoothing Approaches

The USGS uses fixed- and adaptive-kernel (FK and AK) smoothing seismicity approaches (Helmstetter et al., 2007; Moschetti et al., 2015) to forecast background seismicity rates. These approaches use de-clustered catalogs, where dependent earthquakes such as aftershocks have been removed. These approaches only predict epicenters of background earthquakes (up to the appropriate maximum magnitude for the seismotectonic zone; see Moschetti et al., 2015) that are not associated with known rupture areas. Therefore, they would ideally be coupled with a fault model to forecast rupture areas for large earthquakes.

In these approaches, the seismotectonic region is divided into small cells (Figure 3-1). The density of earthquakes in each cell is then calculated by smoothing the location of each earthquake with an isotropic Gaussian kernel, $K_\sigma(\vec{r})$:

$$K_\sigma(\vec{r}) = C(\sigma) \exp\left(\frac{-|\vec{r}|^2}{2\sigma^2}\right)$$

where σ is the smoothing distance and $C(\sigma)$ is a normalizing factor, so that the integral of $K_\sigma(\vec{r})$ over an infinite area is equal to 1:

$$\int K_\sigma(\vec{r}) d\vec{r} = 1$$

Then, the density (μ) at any point \vec{r} is estimated by:

$$\mu(\vec{r}) = \sum_j^N K_{\sigma_j}(\vec{r} - \vec{r}_j).$$

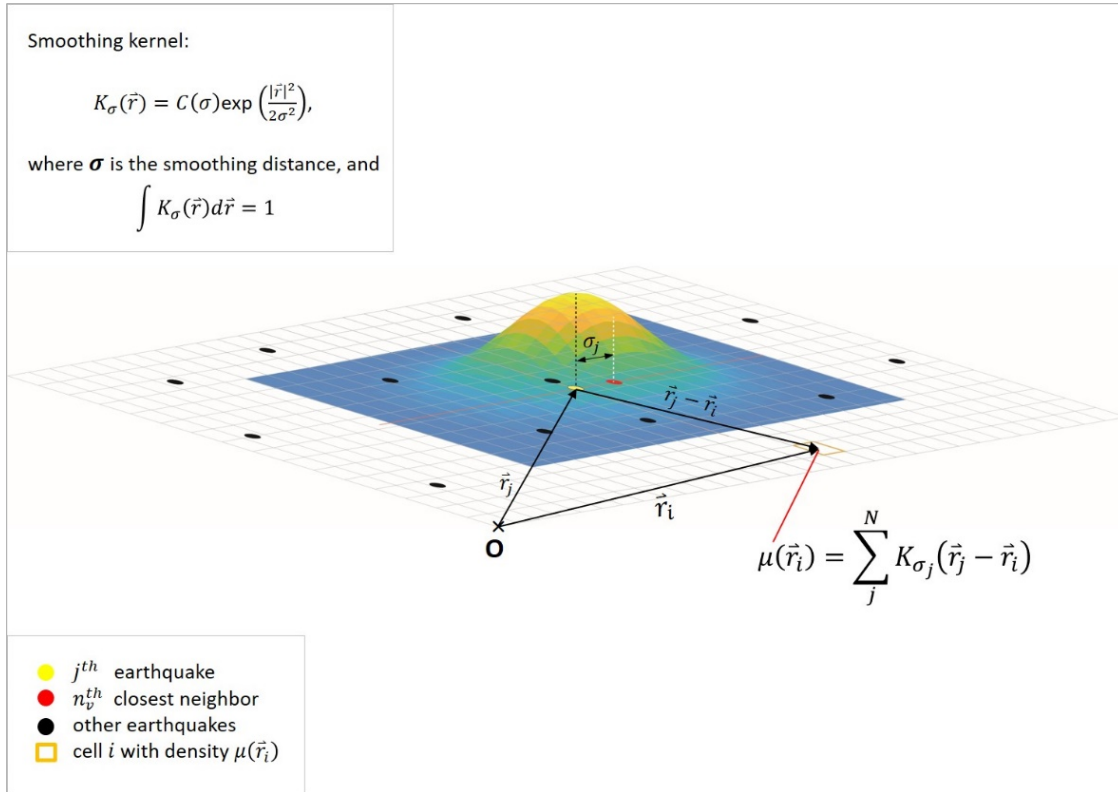


Figure 3-1: Schematic representation of the location of earthquake j smoothed with the isotropic adaptive kernel, $K_{\sigma_j}(\vec{r}_j)$ is shown here. The density of seismicity in each cell is estimated by adding the contributions from all earthquake kernels in that cell.

The adaptive smoothing distance (or kernel bandwidth) σ_j , associated with earthquake j , is computed from the epicentral distances between the event and neighboring events (Helmstetter et al., 2007). Smoothing distance is measured as the epicentral distance between event j and the n_v^{th} closest neighbor. The number of neighbors, n_v , is an adjustable parameter estimated by optimizing the forecasts (likelihood of the model). The model imposes $200 \geq \sigma_j \geq 3.0$ km to account for earthquake location accuracy and grid-cell discretization, with maximum smoothing distances imposed to minimize smoothing of seismicity rates beyond distances of the dimension of potential maximum-magnitude earthquakes. The kernel bandwidth σ_j thus decreases if the density of seismicity at the location of this earthquake (\vec{r}_j) increases. As a consequence, smoothing is “tighter” (smaller σ_j) where the earthquake density is higher.

The fixed and adaptive smoothed seismicity models differ only in the application of the smoothing distance σ_j applied to each earthquake. Fixed smoothed seismicity models use a single smoothing distance for all earthquakes, whereas adaptive smoothed seismicity models use a unique smoothing distance for each earthquake calculated based on the density of the neighboring earthquakes.

Assuming that earthquake occurrence follows a Poisson distribution, the probability of observing n_i events in cell i is calculated from:

$$p(\mu_i^*, n_i) = \frac{(\mu_i^*)^{n_i} e^{-\mu_i^*}}{n_i!}$$

where

$$\mu_i^* = \frac{\mu_i N_{\text{test}}}{\sum_{i=1}^{N_{\text{cell}}} \mu_i}$$

is the rate-normalization that equalizes the number of modeled and observed earthquakes for likelihood testing. Here, the model event density μ_i in cell i is built on a training period of the catalog (input) and tested (evaluated the likelihood) on a separate testing period of the *same* catalog (target). Table 3-1 describes input and target catalog parameters from one of the smoothed seismicity models of Helmstetter et al. (2007). Training and target events are selected from the same catalog, but different time intervals, such that the earthquakes used to estimate μ_i are not the same as the target events to test the model.

Table 3-1: An example of input (training) and target (testing) catalog information used for evaluating model event density μ and log-likelihood LL , respectively (from Helmstetter et al., 2007).

Model Number	Input Catalog				Target Catalog			
	t_1	t_2	M_{\min}	N_{input}	t_1	t_2	M_{\min}	N_{test}
example	1/1/1981	1/1/1996	2.0	52651	1/1/1996	8/23/2005	3.0	2747

The joint log likelihood associated with a smoothed seismicity model m is:

$$\begin{aligned}
LL_m &= \sum_i \ln[p(\mu_i^*, n_i)] \\
&= \sum_i n_i \ln(\mu_i^*) - \sum_i \mu_i^* - \sum_i \ln(n_i!) \\
&\approx \sum_i n_i \ln(\mu_i^*) - \sum_i \mu_i^* - \sum_i [n_i \times \ln(n_i) - n_i + 1] \\
&= \sum_i n_i \ln(\mu_i^*) - N_{\text{test}} - \sum_i n_i \times \ln(n_i) + \sum_i n_i - N_{\text{cell}} \\
&= \sum_i n_i \ln(\mu_i^*) - \sum_i n_i \times \ln(n_i) - N_{\text{test}} + N_{\text{test}} - N_{\text{cell}} \\
&= \sum_i n_i \ln(\mu_i^*) - \sum_i n_i \times \ln(n_i) - N_{\text{cell}}
\end{aligned}$$

Comparisons between smoothed seismicity models use the information-gain parameter G_m , which describes the probability gain per earthquake of using the smoothed seismicity model m compared with the probability of observing the set of earthquake locations predicted by a reference model. The reference model consists of cells with a uniform seismicity rate and has the corresponding joint log-likelihood value LL_{ref} :

$$G_m = \exp\left(\frac{LL_m - LL_{\text{ref}}}{N_{\text{test}}}\right)$$

$$\begin{aligned}
&= \exp\left(\frac{\sum n_i \ln(\mu_i^*) - \sum n_i \ln(\mu_{\text{ref}}^*)}{N_{\text{test}}}\right) \\
&= \exp\left(\frac{\sum n_i \ln(\mu_i^*) - N_{\text{test}} \ln(\mu_{\text{ref}}^*)}{N_{\text{test}}}\right) \\
&= \exp\left(\frac{1}{N_{\text{test}}} \sum_i n_i \ln(\mu_i^*) - \ln(\mu_{\text{ref}}^*)\right) \\
&= \frac{1}{\mu_{\text{ref}}^*} \exp\left(\frac{1}{N_{\text{test}}} \sum_i n_i \ln(\mu_i^*)\right).
\end{aligned}$$

Information gain, therefore, depends only on the seismicity rates in those cells containing earthquakes and is maximized when earthquakes occur in cells with high seismicity rates. Because the joint Poissonian likelihood underlies the calculation of information gain, this optimization parameter is sensitive only to earthquakes that occurred in the testing catalog, and cells containing no earthquakes do not contribute to this calculation (Figure 3-2).

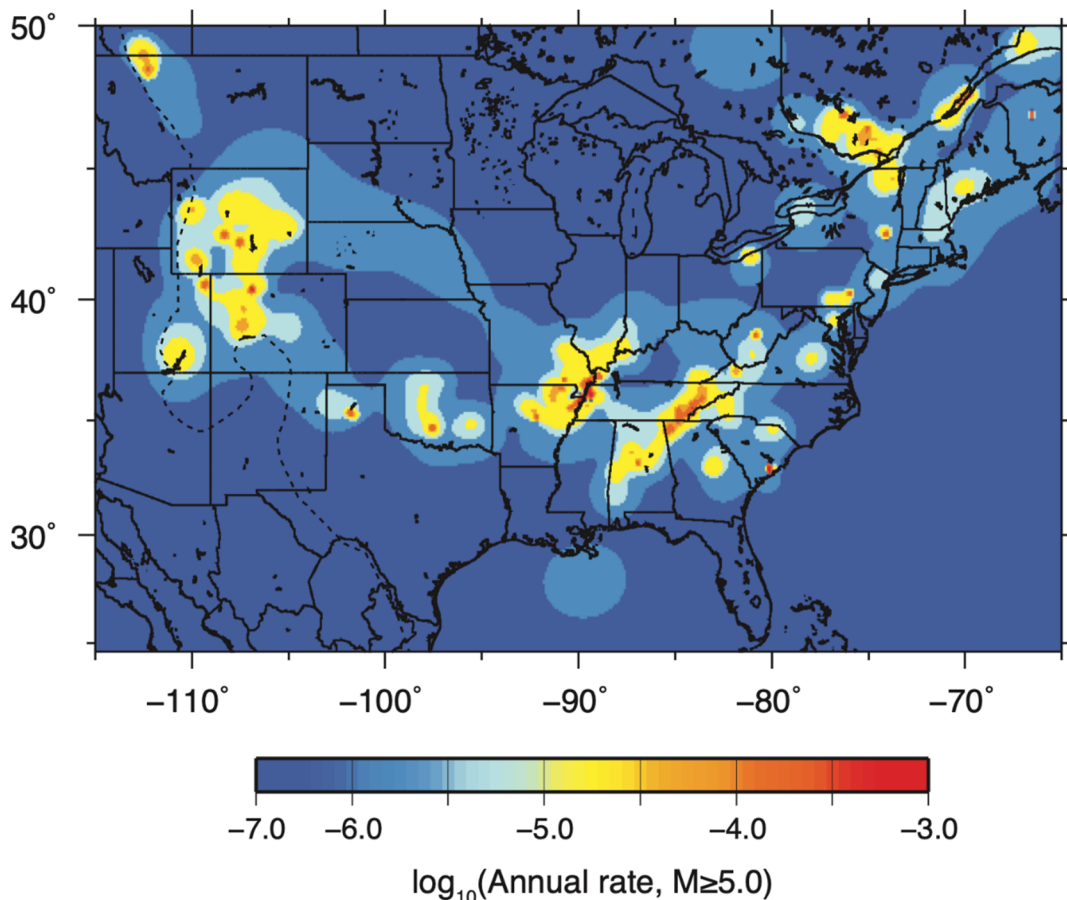


Figure 3-2: Earthquake rate model for the Central and Eastern United States (CEUS) from an adaptive smoothed seismicity model. Annual rates are plotted in a log10 scale and represent the number of $M_w \geq 5.0$ earthquakes within each grid cell. The dashed line between western North America and Central and Eastern North America (CENA) delineates the western boundary of the CEUS for the purposes of this study (Figure 4 in Moschetti, 2015).

For the seismicity-rate and seismic-hazard comparisons of this report, adaptive smoothing models use the fourth nearest neighbor ($n_v = 4$). This value was determined in the calculations for the 2014 NSHM (Petersen et al., 2014). Note that this value differs slightly from the optimized value of Moschetti et al. (2015), which used an earlier version of the earthquake catalog and different methods for computing earthquake completeness times. Recognizing that minor differences in information gain from different neighbor numbers can have substantial effects on spatial smoothing rates. A recent application of adaptive smoothing methods for the Hawaii-NSHM used multiple neighbor numbers to compute the mean seismicity rate forecast (Petersen et al., 2022).

3.2 Penalized-Likelihood Smoothing Approach

The CEUS-SSC uses the penalized-likelihood smoothing approach to estimate recurrence rates of damaging earthquakes ($M_w \geq 5.0$) in the defined seismotectonic zones. The mathematical development of the approach presented in this RIL is based on the formulation described in U.S. Nuclear Regulatory Commission Report Designation 2115, hereafter NUREG-2115 (NRC, 2012).

3.2.1 Likelihood Function

The spatial smoothing operation in this approach is based on calculations of earthquake recurrence within one-quarter-degree or one-half-degree cells, with allowance for “communication” between the cells. Assuming a Gutenberg-Richter magnitude distribution (see equation below), both earthquake rate ν - and β - values are allowed to vary, but the degree of variation has been optimized such that β -values vary little across the study region. Also, ν -value distributions are neither “spiky,” reflecting too strong of a reliance on the exact locations and rate densities of observed events, nor too smooth, reflecting the assumption that the observed record does not provide a spatial constraint on rate density variation. Likewise, the recurrence calculation considers weighting of magnitudes in the recurrence rate calculations; thus, moderate events are assigned more weight than smaller events.

The CEUS SSC model gives strong consideration to locations of repeated large magnitude earthquake (RLME) sources as being spatially stationary through the future time periods of interest, while acknowledging epistemic uncertainties in RLME source locations and geometries. The smoothing operation within the distributed seismicity zones results in variations in ν - and β - values over spatial scales that were judged to be reasonable, given the technical community’s views on spatial stationarity and the relationship between observed small- to moderate-magnitude seismicity and future moderate- to large-magnitude seismicity.

In this section, we describe the penalized-likelihood smoothing approach described NUREG-2115 and introduce and discuss the key formulations of the calculation of earthquake recurrence parameters for one source zone (or one cell within a source zone) under very simple assumptions. The summary below also reformulates and expands the original formulation to include several elements that make the formulation more robust, realistic, and flexible. These elements include the reformulation in terms of magnitude bins, and the introduction of magnitude-dependent weights, catalog incompleteness, the effect of maximum earthquake magnitude (M_{\max}), spatial variation of parameters within the source zone, and the prior distributions of β .

The penalized-likelihood model assumes a Gutenberg-Richter distribution. Therefore, the number of earthquakes, $N(m)$, with magnitudes equal to m or larger is:

$$N(m) = AT\nu e^{-\beta(m-m_0)} \quad m \geq m_0 \quad 3-1$$

where A is the source area, T the duration of time window, and ν the earthquake rate (per unit area, per unit time).

Then, the number of earthquakes (n) with magnitudes in the interval between m and $m + dm$ that occur in a source zone during time T is given by

$$n = -\frac{dN(m)}{dm} = AT\nu\beta e^{-\beta(m-m_0)}$$

Assuming that n follows a Poisson probability distribution ($p(k, n) = n^k \frac{e^{-n}}{k!}$), the probability distribution for $k = 1$ (i.e., either 1 or 0 event) takes the following form:

$$p(1, n) = AT\nu\beta e^{-\beta(m-m_0)} \times e^{-[AT\nu\beta e^{-\beta(m-m_0)}]}$$

Then, the likelihood function for a total of N earthquakes in the catalog becomes

$$l(\nu, \beta) = \left[\prod_{i=1}^N AT\nu\beta e^{-\beta(m_i-m_0)} \right] \times e^{-\sum AT\nu\beta e^{-\beta(m_i-m_0)}}$$

Assuming a continuous magnitude distribution,

$$l(\nu, \beta) = \left[\prod_{i=1}^N AT\nu\beta e^{-\beta(m_i-m_0)} \right] \times \exp\left(-A\nu T \int_{m_0}^{\infty} \beta e^{-\beta(m-m_0)} dm\right) \quad 3-2$$

and because

$$\int_{m_0}^{\infty} \beta e^{-\beta(m-m_0)} dm = 1,$$

the likelihood function reduces to the following:

$$l(\nu, \beta) = (AT\nu)^N e^{-A\nu T} \prod_{i=1}^N (\beta e^{-\beta(m_i-m_0)}) \quad 3-3$$

At small magnitudes, the data deviate substantially from a smooth exponential shape (Figure 3-3). These deviations are not consistent with the exponential model, given the very large earthquake counts in these bins. The large earthquake counts mean that the continuous formulation of the likelihood may incorrectly interpret the local slope of the histogram around magnitude 3.

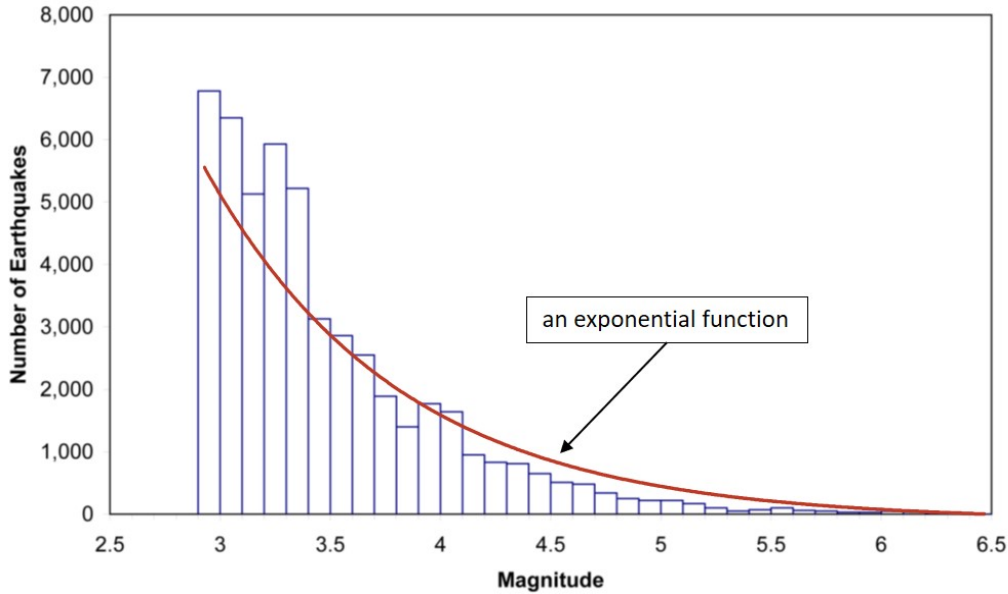


Figure 3-3: Histogram of magnitudes in the earthquake catalog used in Central and Eastern United States Seismic Source Characterization (CEUS-SSC). The minimum magnitude shown (M_W 2.9) is the lowest magnitude used in these recurrence calculations (Figure 5.3.2-3, NUREG-2115, v2; NRC 2012). The red exponential curve highlights deviation of the data from a smooth exponential shape below M_W 3.5.

To avoid these potential problems, several modifications are made to the likelihood function that are described below.

Modification Number 1

The likelihood function is reformulated in terms of the earthquake counts in discrete magnitude bins, with bin sizes appropriately selected so that magnitudes converted from intensity fall in the middle of these bins. Considering bins of δ magnitude units (i.e., first bin between magnitudes m_0 and $m_0 + \delta$, second bin between $m_0 + \delta$ and $m_0 + 2\delta$, etc.),

$$\left| \begin{array}{c} m^1_0, n_1 \\ m_0 \qquad m_0 + \delta \end{array} \right| \left| \begin{array}{c} m^2_0, n_2 \\ m_0 + \delta \qquad m_0 + 2\delta \end{array} \right| \dots \left| \begin{array}{c} m^k_0, n_k \\ m_0 + k\delta \qquad m_0 + (k+1)\delta \end{array} \right| \dots$$

one can write the likelihood function as follows:

$$l(v, \beta) = l_1 \times l_2 \dots \times l_k \dots$$

where

$$\begin{aligned}
l_k(\nu, \beta) &= \left(\prod_{i=1}^{n_k} AT\nu\beta e^{-\beta(m_i^k - m_0^k)} \right) \times \exp \left(- \sum_i^{n_k} AT\nu\beta e^{-\beta(m_i^k - m_0^k)} \right) \\
&= \prod_{i=1}^{n_k} AT\nu p_k \times \exp \left(- \sum_i^{n_k} AT\nu p_k \right) \\
&= (AT\nu)^{n_k} \prod_{i=1}^{n_k} p_k \times \exp \left(-Av \sum_i^{n_k} T p_k \right)
\end{aligned}$$

Here n_k is the number of earthquakes in the k^{th} magnitude bin $m_0^k \leq m < m_0^{k+1}$, and

$$\begin{aligned}
p_k &= \beta e^{-\beta(m_i^k - m_0^k)} \\
&= \beta e^{-\beta(m_i^k - m_0)} - \beta e^{-\beta(m_i^{k+1} - m_0)}.
\end{aligned}$$

Then,

$$l(\nu, \beta) = \prod_{k=1}^{\infty} \left\{ (AT\nu)^{n_k} \prod_{i=1}^{n_k} p_k \times \exp \left(-Av \sum_i^{n_k} T p_k \right) \right\},$$

or

$$l(\nu, \beta) = \prod_{k=1}^{\infty} \left\{ (AT\nu)^{n_k} \prod_{i=1}^{n_k} p_k \right\} \times \exp \left(-Av \sum_{k=1}^{\infty} T p_k \right). \quad 3-4$$

Note: The number n_k in Equation 3-4 is not necessarily an integer number. The effect of uncertainty in magnitude is accounted by calculating an “event factor” for each earthquake (see Chapter 3, NUREG-2115, v2; NRC 2012). Thus, n_k represents the sum of the event factors for all the earthquakes in the k^{th} magnitude bin.

Modification Number 2

A second modification is required because the assessment may result in a lower weight for lower magnitude bins. For example, the magnitude-recurrence law may deviate from exponential, or the magnitude-conversion models or completeness model may be less reliable for lower magnitudes. One can represent these weights by the quantities, $w_1, w_2, \dots, w_k, \dots$, indexed by k , the magnitude-bin number; note also that these weights do not necessarily add to one, and the resulting weighted likelihood function is:

$$l(\nu, \beta) = \prod_{k=1}^{\infty} (AT\nu p_k)^{n_k w_k} \times \exp \left(-Av \sum_{k=1}^{\infty} T p_k w_k \right). \quad 3-5$$

Modification Number 3

A third modification is necessary to account for magnitude-dependent catalog completeness. Following EPRI-SOG (1986), earthquakes in magnitude bin k are counted if they occurred after a certain date $t_{0,k}$, even if the catalog is not fully complete after that date, and the associated catalog duration is characterized by the *Equivalent Period of Completeness*, which is defined as

$$T_{E,k} = \int_{t_{0,k}}^{\text{present}} P_D(t, k) dt \quad 3-6$$

where $P_D(t, k)$ is the detection probability for magnitudes in bin k at time t . After introducing magnitude-dependent completeness, the likelihood function becomes

$$l(\nu, \beta) = \prod_{k=1}^{\infty} (AT_{E,k} \nu p_k)^{n_k w_k} \times \exp\left(-A\nu \sum_{k=1}^{\infty} T_{E,k} p_k w_k\right). \quad 3-7$$

Modification Number 4

A fourth modification is necessary to account for maximum magnitude M_{\max} . In the case of a single M_{\max} value, p_k is modified by considering that portions of a magnitude bin, or the entire bin, may be above M_{\max} , resulting in the expression:

$$p_k(M_{\max}) = \begin{cases} \exp[-\beta(m_0^k - m_0)] - \exp[-\beta(m_0^{k+1} - m_0)], & m_0^{k+1} < M_{\max} \\ \exp[-\beta(m_0^k - m_0)] - \exp[-\beta(M_{\max} - m_0)], & m_0^k < M_{\max} < m_0^{k+1} \\ 0, & M_{\max} < m_0^k \end{cases} \quad 3-8$$

The normalizing constant $(1 - \exp[-\beta(M_{\max} - m_0)])$ in the above equation is neglected because it is nearly equal to unity ($m_0 = 2.9$ and $M_{\max} > 5.5$).

For the case where there is epistemic uncertainty in M_{\max} (which is represented in this project by a five-point discrete distribution: 6.1, 6.7, 7.2, 7.7, 8.1), one can use the *expected* value (with respect to M_{\max}) of the p_k given by Equation 3-8; that is,

$$p_k = E_{M_{\max}}[p_k(M_{\max})]. \quad 3-9$$

The first step in the penalized-likelihood approach is to divide the source zones into cells for the purposes of the recurrence-model evaluation and hazard calculations. This division is often done along parallels and meridians. The CEUS-SSC uses a cell dimension of one-quarter degree by one-quarter degree. For source zones of large dimensions, the cell size is increased to one-half degree by one-half degree for computational efficiency. Cells near source-zone boundaries have smaller areas and irregular shapes, generated so that the geometry of a source zone is honored in the recurrence and hazard calculations, without any quantization error.

The likelihood function is then formulated separately for each cell in the source zone. The joint likelihood function for the values of ν and β in all M cells within the source zone takes the form:

$$l(\mathbf{v}, \boldsymbol{\beta}) = \prod_{j=1}^M \left\{ \prod_{k=1}^Q \left(v_j A_j T_E(k, j) p(k, j) \right)^{n(k, j) w_k} \exp \left(-v_j A_j \sum_{k=1}^Q T_E(k, j) p(k, j) w_k \right) \right\}. \quad 3-10$$

Q denotes the highest bin with non-zero p_k (because M_{\max} is finite, the number of magnitude bins becomes finite). Vectors \mathbf{v} , $\boldsymbol{\beta}$ contain the values of v and β in all cells within the source zone, j is the index for the cells ($j = 1, \dots, M$). Also, the notation has been changed slightly from previous equations, so that $T_E(k, j)$, $n(k, j)$, and $p(k, j)$ are the equivalent period of completeness, earthquake count, and truncated-exponential bin probability (Equation 3-9) for magnitude bin k and cell j .

3.2.2 Penalty Function

Because the cells are small and most do not contain enough earthquakes to allow a reliable estimation of v_j and β_j using earthquakes alone, and because very large variations in v and β (particularly the latter) between adjacent cells are not considered physically realistic, one introduces penalty functions that “penalize” those solutions where v and β have large variations between adjacent cells. Thus, the solution to the penalized-likelihood problem represents an optimal compromise between consistency with the data (as indicated by a high value of the likelihood) and smoothness (as indicated by a low value of the penalty function).

The penalty function is based on the Laplacian operator, which represents a deviation from local average. For a function $f(x, y)$ in two dimensions (Figure 3-4), the Laplacian operator and its lowest-order finite difference approximation are given by the following:

$$\begin{aligned} \Delta f(x, y) &= \nabla^2 f(x, y) \\ &= \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \approx \frac{\Delta^2 f}{\Delta x^2} + \frac{\Delta^2 f}{\Delta y^2} \\ &= 2 \left(\frac{f(x - \Delta x, y)/2 + f(x + \Delta x, y)/2 - f(x, y)}{\Delta x^2} \right) \\ &\quad + 2 \left(\frac{f(x, y - \Delta y)/2 + f(x, y + \Delta y)/2 - f(x, y)}{\Delta y^2} \right) \\ &= -2 \left(\frac{f(x, y) - \bar{f}^x(x, y)}{\Delta x^2} \right) - 2 \left(\frac{f(x, y) - \bar{f}^y(x, y)}{\Delta y^2} \right) \end{aligned} \quad 3-11$$

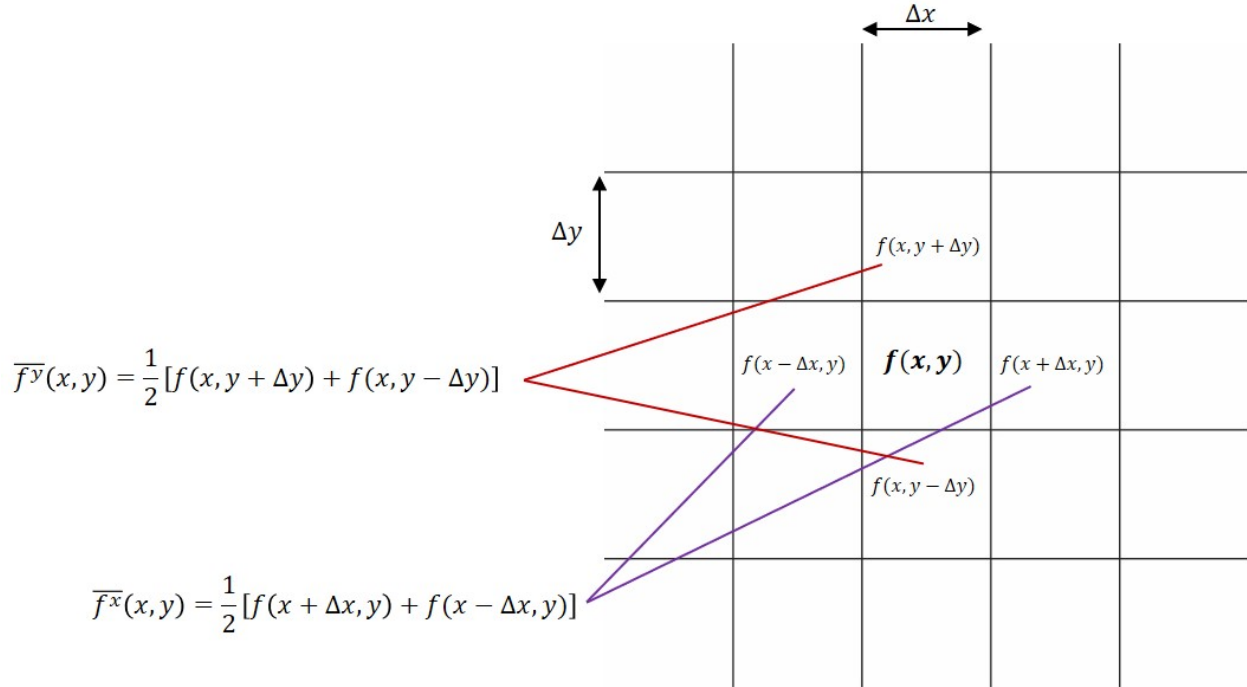


Figure 3-4: A schematic representation of $f(x, y)$ in a cell, and in its four neighbors.

Based on the above, and assuming a *normal distribution* for variations in β between a cell and its neighbors, the penalty function for β over the entire source zone is of the form

$$f_{\beta}(\mathbf{\beta}, \sigma_{\Delta\beta}) = \prod_{j=1}^M \frac{1}{\sqrt{2\pi}\sigma_{\Delta\beta}} \exp \left[-\frac{1}{2} \left(\frac{\Delta\beta_j}{\sigma_{\Delta\beta}} \right)^2 \right]. \quad 3-12$$

where the product extends over all cells; $\sigma_{\Delta\beta}$ controls the degree of smoothing, and $\Delta\beta_j$ is given by (Figure 3-5):

$$\Delta\beta_j = -\frac{1}{2} \left[\frac{\beta(x_j, y_j) - \bar{\beta}^x(x_j, y_j)}{\Delta x^2} \right] - \frac{1}{2} \left[-\frac{\beta(x_j, y_j) - \bar{\beta}^y(x_j, y_j)}{\Delta y^2} \right]. \quad 3-13$$

The penalty term $f_v(\mathbf{v}, \sigma_{\Delta v})$ for the rate is constructed in the same manner, except that the Laplacian term Δv_j is calculated in terms of $\ln v$ instead of v itself.

$$f_v(\mathbf{v}, \sigma_{\Delta v}) = \prod_{j=1}^M \frac{1}{\sqrt{2\pi}\sigma_{\Delta v}} \exp \left[-\frac{1}{2} \left(\frac{\Delta(\ln v_j)}{\sigma_{\Delta v}} \right)^2 \right]. \quad 3-14$$

Remark: $\sigma_{\Delta v}$ is actually $\sigma_{\Delta \ln v}$. However, to be consistent with NUREG-2115, the notation is not altered here.

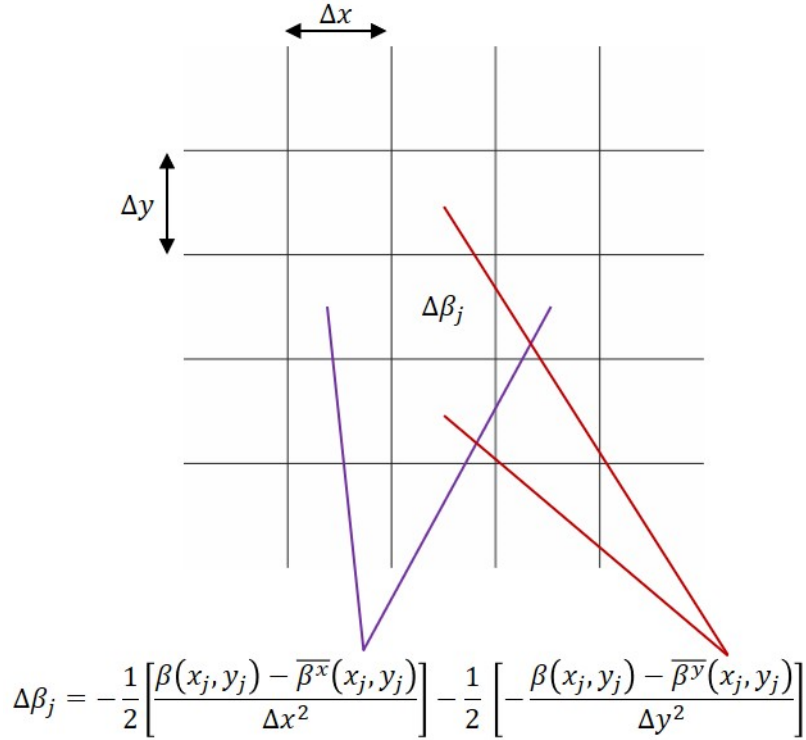


Figure 3-5: A schematic representation of $\beta(x, y)$ deviation from the local average.

There is also an additional penalty term in the form of a prior distribution for β_j . The product of the prior distributions for all cells in the source zone takes the following form:

$$f'_{\beta}(\boldsymbol{\beta}, \sigma_{\beta}) = \prod_{j=1}^M \frac{1}{\sqrt{2\pi}\sigma_{\beta}} \exp \left[-\frac{1}{2} \left(\frac{\beta_j - \beta_{\text{prior}}}{\sigma_{\beta}} \right)^2 \right]. \quad 3-15$$

There is a subtle difference between Equations 3-14 and 3-15. The former favors solutions with spatially uniform β , regardless of the value of β , whereas the latter promotes values of β near β_{prior} .

The complete penalized-likelihood is constructed as the product of the joint likelihood for all the cells in the source zone (Equation 3-10), the smoothness penalty terms (Equations 3-12 and 3-14), and the prior distribution of β (Equation 3-15):

$$p(\mathbf{X}) = l(\mathbf{v}, \boldsymbol{\beta}) f_{\mathbf{v}}(\mathbf{v}, \sigma_{\Delta\mathbf{v}}) f_{\beta}(\boldsymbol{\beta}, \sigma_{\Delta\beta}) f'_{\beta}(\boldsymbol{\beta}, \sigma_{\beta}), \quad 3-16$$

where

$$\mathbf{X} = [\mathbf{v}, \boldsymbol{\beta}, \sigma_{\Delta\mathbf{v}}, \sigma_{\Delta\beta}] \quad 3-17$$

is the complete vector of parameters and hyper-parameters (Figure 3-6).

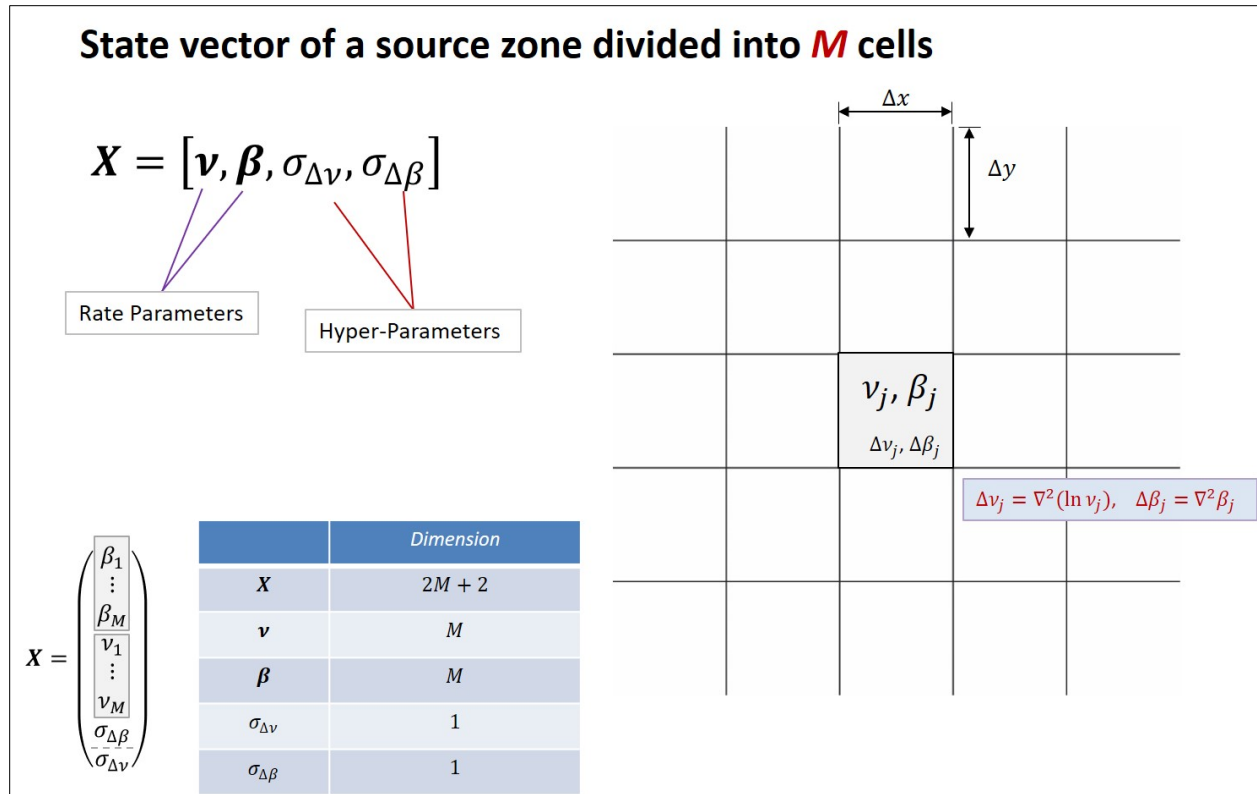


Figure 3-6: State vector, \mathbf{X} , of a source zone divided into M cells.

3.3 Modeling the Joint Distribution of Recurrence Parameters

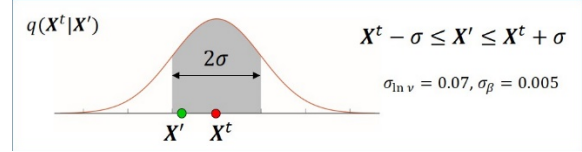
To generate alternative maps of the recurrence parameters, the Metropolis algorithm of Markov Chain Monte Carlo (MCMC) is used. MCMC is frequently used to generate multiple realizations from a complex multi-dimensional probability distribution by constructing a Markov chain that has this distribution as its limiting or stationary distribution (Equation 3-16 times a constant). This distribution represents the central tendency and epistemic uncertainty of $\mathbf{X} = [\mathbf{v}, \boldsymbol{\beta}, \sigma_{\Delta v}, \sigma_{\Delta \beta}]$, which contains the recurrence parameters for all cells (\mathbf{v} and $\boldsymbol{\beta}$), plus the two hyperparameters ($\sigma_{\Delta v}$ and $\sigma_{\Delta \beta}$). The number of dimensions of \mathbf{X} is $2M + 2$, where M is the number of cells in the source zone (Figure 3-6).

A Markov chain is a discrete-time probabilistic model with states $\mathbf{X}(1), \mathbf{X}(2), \dots$, in which the conditional probability distribution of the state at time $t + 1$ (denoted \mathbf{X}^{t+1}), given the states at earlier times, depends *only* on the immediately previous state (\mathbf{X}^t). This conditional probability distribution of \mathbf{X}^{t+1} given \mathbf{X}^t is known as the *transition probability*. If a Markov chain meets certain requirements, it possesses a limiting or stationary distribution, which will be reached asymptotically after many realizations, regardless of the initial state.

MCMC constructs a discrete-time Markov chain that has a stationary or limiting distribution, equal to the joint distribution of interest (i.e., the penalized likelihood function $p(\mathbf{X})$ given by Equation 3-16). One can use this sequence to generate many realizations (~20 million) from that distribution. A portion of the initial realizations (about 25%) are discarded because they are affected by the initial conditions. The remaining realizations are then used to calculate statistics.

In the Metropolis MCMC algorithm, a new realization \mathbf{X}^{t+1} is generated from \mathbf{X}^t by the following two-step procedure:

1. Generate a new candidate state \mathbf{X}' by drawing from a trial distribution $q(\mathbf{X}'|\mathbf{X}^t)$ that depends only on \mathbf{X}^t and is symmetrical, *i.e.*, $q(\mathbf{X}'|\mathbf{X}^t) = q(\mathbf{X}^t|\mathbf{X}')$.



2. Accept the new trial state with probability:

$$P_{\text{accept}} = \min\left(1, \frac{p(\mathbf{X}')}{p(\mathbf{X}^t)}\right) \quad 3-18$$

If the new state is accepted (*i.e.*, if a standard uniform random number generator draws a value lower than P_{accept}), \mathbf{X}' becomes the new state \mathbf{X}^{t+1} . Otherwise, the new state \mathbf{X}^{t+1} is equal to \mathbf{X}^t . The Metropolis algorithm is implemented locally, in the following order:

1. For each cell j , trial values of $\ln v_j$ and β_j are drawn from uniform distributions centered at the current values of these parameters.
2. The penalized likelihood function $p(\mathbf{X})$ (Equation 3-16) is calculated for these trial values.
3. The values are accepted or rejected using Equation 3-18 and a random-number generator. (Accept \mathbf{X}' as the new state if $\ln\left(\frac{p(\mathbf{X}')}{p(\mathbf{X}^t)}\right) > 0$, or greater than the natural log of a randomly generated number between 0 and 1.)
4. The only portions of the penalized likelihood that are reevaluated each time are the likelihood function for cell j , the penalty terms for cell j and for its neighboring cells within the source zone, and the prior distribution of β_j .
5. After going through all cells, trial values of the hyperparameters $\sigma_{\Delta v}$ and $\sigma_{\Delta \beta}$ are drawn in a similar manner (steps 1-4, above), the penalized likelihood is recalculated, and the values are accepted or rejected. In this step, only the smoothness-penalty terms for all cells are recalculated. This process is repeated many times ($\sim 2 \times 10^7$).

The widths of the uniform distributions (in Step 1 above) are controlled by the analyst. These widths affect the acceptance rate, which in turn affects the numerical efficiency of the algorithm. Recommendations as to the optimal acceptance rates vary; a common recommendation is values between 20% and 40%. After discarding some of the initial realizations of \mathbf{X}^t , one generates many realizations to represent the joint distribution of the state vector $\mathbf{X} = [\mathbf{v}, \boldsymbol{\beta}, \sigma_{\Delta v}, \sigma_{\Delta \beta}]$. The section that follows describes the approach to obtain eight equally likely realizations of \mathbf{X} (*i.e.*, eight alternative maps to represent the joint distribution of \mathbf{v} and $\boldsymbol{\beta}$, and the two hyperparameters $\sigma_{\Delta v}$ and $\sigma_{\Delta \beta}$), including their correlation structure.

It is worth noting that a fixed width of $\ln v$ distribution affects the range of acceptable new realizations of the recurrence rate, v , differently. As shown in Figure 3-7, in very low-seismicity regions, the new values tend to remain very close to the previous values; whereas, in more seismically active regions, the new realizations can have large variations.

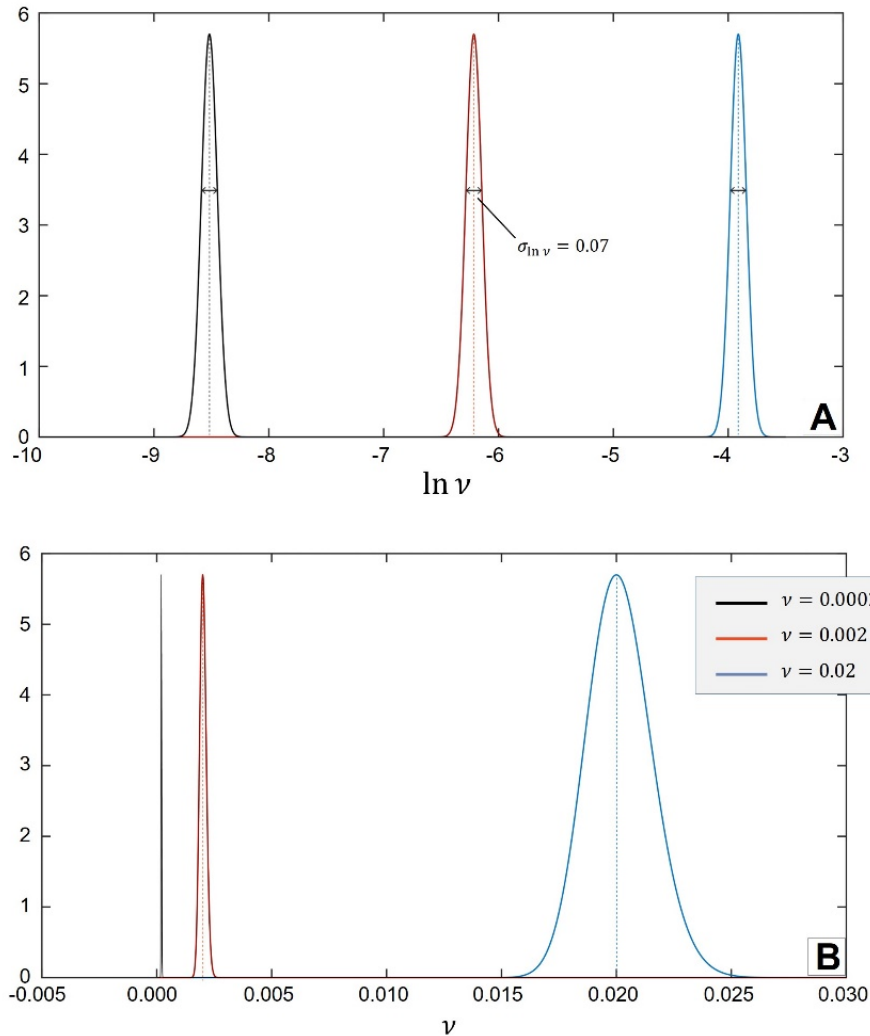


Figure 3-7: Plots of P -acceptance distribution as a function of $\ln \nu$, and ν are shown here. Frame A shows normal distributions of $\ln \nu$ for mean values of 0.02, 0.002, and 0.0002. A standard deviation of 0.07 results in identical shapes for all three distributions. However, the same probabilities plotted as a function of recurrence rate ν (frame B) have a different shape. These plots indicate that, with a fixed standard deviation for Markov Chain Monte Carlo (MCMC) trial distributions to generate a new state candidate, at regions with very low earthquake recurrence rates, the new realizations remain very close to the initial value. In contrast, where the recurrence rate is high, the new MCMC candidate could have a larger deviation from the current value.

3.3.1 Development of Recurrence Maps

Distributed seismicity recurrence rate calculations are performed for each of the designated seismotectonic zones in the CEUS using the MCMC technique. The MCMC formulation, as initial input, uses the recurrence rates in that zone, estimated from the earthquake catalog, to generate many realizations (typically 20 million), of which roughly the first 25% are discarded to minimize potential bias caused by the initial conditions. Of note, all dependent earthquakes are removed

from the catalog to meet the Poisson distribution criterion; all earthquakes associated with the RLME sources also are removed from the catalog to avoid double-counting.

In this report, estimates of earthquake recurrence rates were calculated for seismotectonic zones of North Appalachian (NAP), Paleozoic Extended Crust-Wide (PEZ-W), and Non-Mesozoic-and-Younger Extension-Narrow (NMESE-N), as shown in Figure 3-8.

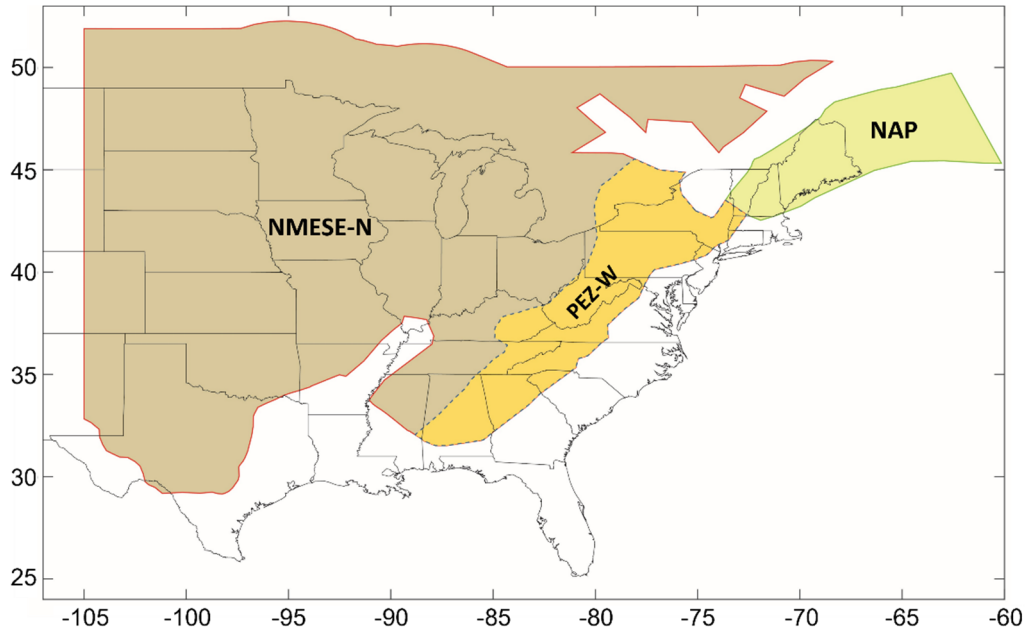


Figure 3-8: The North Appalachian (NAP), Non-Mesozoic-and-Younger Extension-Narrow (NMESE-N), and Paleozoic Extended Crust-Wide (PEZ-W) seismotectonic zones for which the recurrence rates were calculated. Note that PEZ-W is a sub-zone of NMESE-N.

3.3.1.1 Mean Recurrence Maps

As an example, maps of the mean recurrence parameters ν and b for the North Appalachian seismotectonic zone (NAP) are calculated from the average of many of their MCMC realizations in each cell. Figure 3-9 shows maps of the mean recurrence rate, ν , (for earthquakes larger than $M5$) and the b -value (Gutenberg–Richter slope). For reference, the seismic activity in the NAP zone from the 2018 USGS catalog also is shown. As expected, the estimated recurrence rate map correlates well with the seismicity map and the b -value does not deviate significantly from unity in the zone.

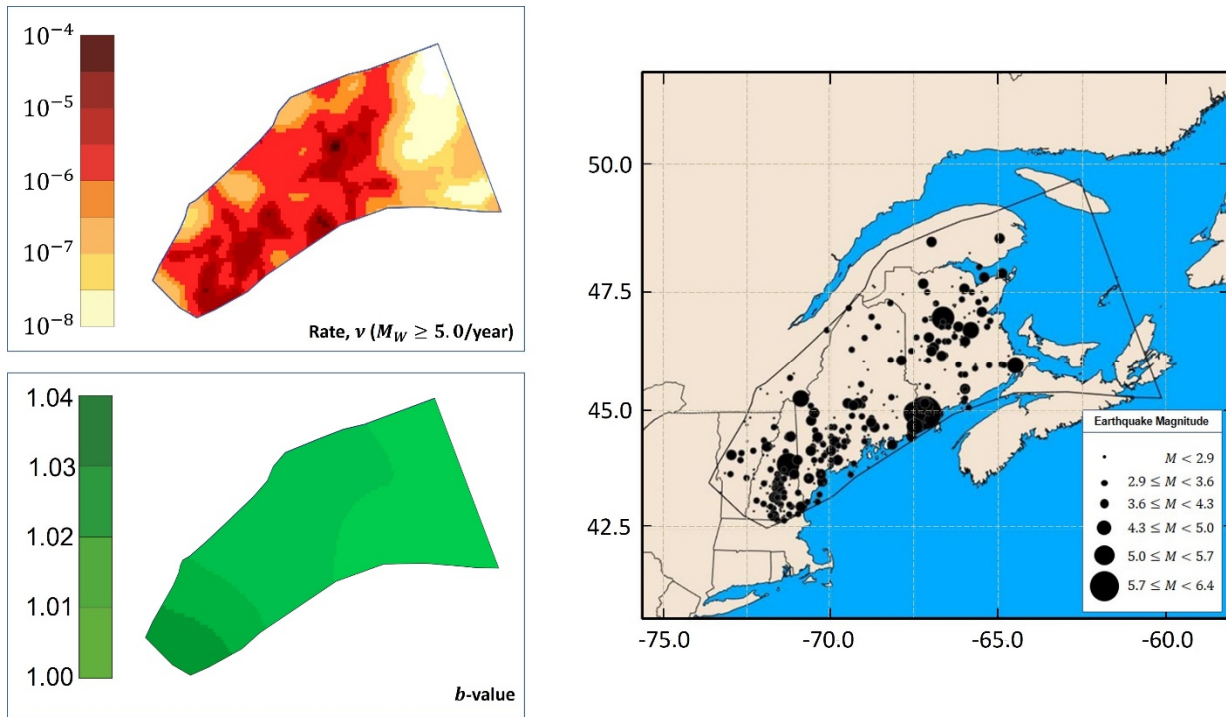


Figure 3-9: Maps of the mean recurrence rate ($M_W \geq 5.0$ events/year) and the b -values for the North Appalachian (NAP) seismotectonic zone estimated by using the Penalized-Likelihood smoothing method are shown on the left column. The grid size in this case is $0.1^\circ \times 0.1^\circ$.

3.3.2 Alternative Recurrence Maps

To represent the uncertainty in recurrence parameters ν and β that results from the limited duration of the catalog, we construct several alternative maps that jointly represent the central tendency and statistical uncertainty in these parameters, including correlations between the recurrence parameters of cells in the same geographical region. The steps presented below show how these maps are constructed.

Step 1: Generate many realizations of the state vector $X [\nu, \beta, \sigma_{\Delta\nu}, \sigma_{\Delta\beta}]$ using the MCMC algorithm described above. Use these realizations to calculate the mean vector m_X and the covariance matrix S_X of X . The covariance matrix S_X (Figure 3-10) contains information about the marginal variance of ν and β in each cell, the correlation between ν and β in each cell, and the correlation between parameters in neighboring cells.

$$\mathbf{X} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_M \\ v_1 \\ \vdots \\ v_M \\ \sigma_{\Delta v} \\ \sigma_{\Delta\beta} \end{bmatrix}, \quad \mathbf{X} - \mathbf{m}_X = \begin{bmatrix} \beta_1 - \bar{\beta} \\ \vdots \\ \beta_M - \bar{\beta} \\ v_1 - \bar{\alpha} \\ \vdots \\ v_M - \bar{\alpha} \\ \sigma_{\Delta v} - \bar{\sigma}_{\Delta v} \\ \sigma_{\Delta\beta} - \bar{\sigma}_{\Delta\beta} \end{bmatrix}, \quad \mathbf{S}_X = \overline{(\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^T}$$

j \ i	1	2	...	M+1	M+2	...	2M	2M+1	2M+2
1	$\overline{\beta_1\beta_1} - \bar{\beta}_1\bar{\beta}_1$								
2	$\overline{\beta_2\beta_1} - \bar{\beta}_2\bar{\beta}_1$	$\overline{\beta_2\beta_2} - \bar{\beta}_2\bar{\beta}_2$							
⋮			⋮			⋮			
M+1	$\overline{v_1\beta_1} - \bar{v}_1\bar{\beta}_1$	$\overline{v_1\beta_2} - \bar{v}_1\bar{\beta}_2$		$\overline{v_1v_1} - \bar{v}_1\bar{v}_1$					
M+2	$\overline{v_2\beta_1} - \bar{v}_2\bar{\beta}_1$	$\overline{v_2\beta_2} - \bar{v}_2\bar{\beta}_2$		$\overline{v_2v_1} - \bar{v}_2\bar{v}_1$	$\overline{v_2v_2} - \bar{v}_2\bar{v}_2$				
⋮			⋮			⋮			
2M	$\overline{v_{2M}\beta_1} - \bar{v}_{2M}\bar{\beta}_1$	$\overline{v_{2M}\beta_2} - \bar{v}_{2M}\bar{\beta}_2$		$\overline{v_{2M}v_1} - \bar{v}_{2M}\bar{v}_1$	$\overline{v_{2M}v_2} - \bar{v}_{2M}\bar{v}_2$		$\overline{v_{2M}v_{2M}} - \bar{v}_{2M}\bar{v}_{2M}$		
M+1	$\overline{\sigma_{\Delta v}\beta_1} - \bar{\sigma}_{\Delta v}\bar{\beta}_1$	$\overline{\sigma_{\Delta v}\beta_2} - \bar{\sigma}_{\Delta v}\bar{\beta}_2$		$\overline{\sigma_{\Delta v}v_1} - \bar{\sigma}_{\Delta v}\bar{v}_1$	$\overline{\sigma_{\Delta v}v_2} - \bar{\sigma}_{\Delta v}\bar{v}_2$		$\overline{\sigma_{\Delta v}v_{2M}} - \bar{\sigma}_{\Delta v}\bar{v}_{2M}$	$\overline{\sigma_{\Delta v}\sigma_{\Delta v}} - \bar{\sigma}_{\Delta v}\bar{\sigma}_{\Delta v}$	
2M+2	$\overline{\sigma_{\Delta\beta}\beta_1} - \bar{\sigma}_{\Delta\beta}\bar{\beta}_1$	$\overline{\sigma_{\Delta\beta}\beta_2} - \bar{\sigma}_{\Delta\beta}\bar{\beta}_2$		$\overline{\sigma_{\Delta\beta}v_1} - \bar{\sigma}_{\Delta\beta}\bar{v}_1$	$\overline{\sigma_{\Delta\beta}v_2} - \bar{\sigma}_{\Delta\beta}\bar{v}_2$		$\overline{\sigma_{\Delta\beta}v_{2M}} - \bar{\sigma}_{\Delta\beta}\bar{v}_{2M}$	$\overline{\sigma_{\Delta\beta}\sigma_{\Delta v}} - \bar{\sigma}_{\Delta\beta}\bar{\sigma}_{\Delta v}$	$\overline{\sigma_{\Delta\beta}\sigma_{\Delta\beta}} - \bar{\sigma}_{\Delta\beta}\bar{\sigma}_{\Delta\beta}$

Figure 3-10: Covariance matrix of a source zone divided into M cells. Due to symmetry, only the lower triangle (covariance) and the trace (variance) are shown here.

Step 2: Perform an eigenvalue decomposition of the covariance matrix \mathbf{S}_X . Because this matrix is a covariance matrix, it is positive-semi-definite. Therefore, all eigenvalues are positive or zero. Let $\alpha_1^2, \alpha_2^2, \dots, \alpha_{2M+2}^2$ be the eigenvalues of \mathbf{S}_X , ordered from largest to smallest, and let $\phi_1, \phi_2, \dots, \phi_{2M+2}$ be the associated eigenvectors. An important property of this eigenvalue decomposition is that of orthogonality; that is, the dot product of ϕ_r and ϕ_s is equal to 1 if $r = s$ and 0 if $r \neq s$. Therefore, a *random* vector $\tilde{\mathbf{X}}$ can be constructed using the following randomized linear combination of the eigenvectors:

$$\tilde{\mathbf{X}} = \mathbf{m}_X + \sum_{k=1}^{2M+2} \varepsilon_k \phi_k \quad 3-19$$

where $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{2M+2}$ are independent random variables with mean zero and variances $\alpha_1^2, \alpha_2^2, \dots, \alpha_{2M+2}^2$. This implies that a synthesized realization of $\tilde{\mathbf{X}}$ generated using Equation 3-19 has the correlation properties of the random vector $\mathbf{X} [\mathbf{v}, \boldsymbol{\beta}, \sigma_{\Delta v}, \sigma_{\Delta\beta}]$ that we want to simulate.

$$X = [v, \beta, \sigma_{\Delta v}, \sigma_{\Delta \beta}]$$

$$S_X = \overline{(X - m_X)(X - m_X)^T}$$

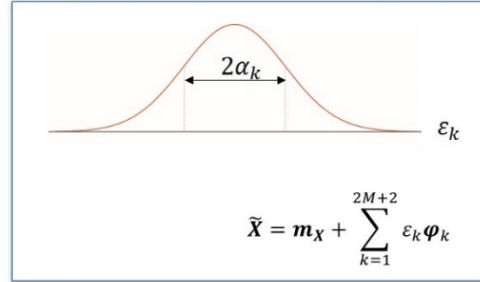
Compute eigenvalues S_X through Singular Value Decomposition (SVD):

$$S_X = \Phi \Sigma \Psi^T$$

$$\Phi = \begin{pmatrix} \varphi_{1,1} & \cdots & \varphi_{1,2M+2} \\ \vdots & \ddots & \vdots \\ \varphi_{2M+2,1} & \cdots & \varphi_{2M+2,2M+2} \end{pmatrix}$$

$$\Psi = \begin{pmatrix} \psi_{1,1} & \cdots & \psi_{1,2M+2} \\ \vdots & \ddots & \vdots \\ \psi_{2M+2,1} & \cdots & \psi_{2M+2,2M+2} \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \alpha_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha_{2M+2}^2 \end{pmatrix}$$



In many applications of this technique, only a few terms are required in Equation 3-19 because the eigenvalues decrease rapidly in size. This is not the case in this application, and it may be necessary to use a large number of terms (up to a few hundred, e.g., $k \sim 400$) in the summation in order to represent 99% of the total variance of X (Figure 3-11).

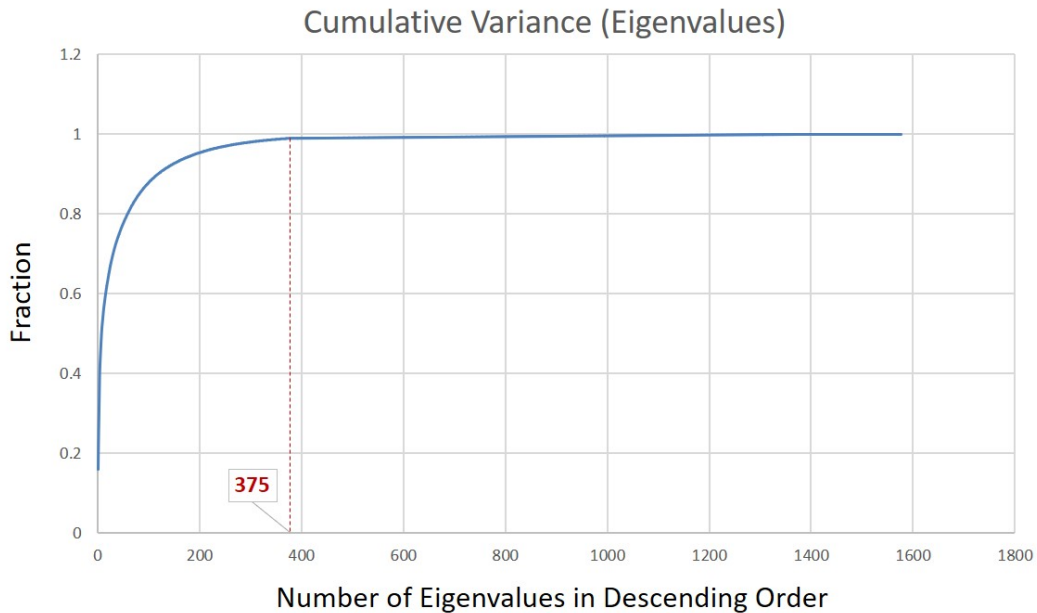
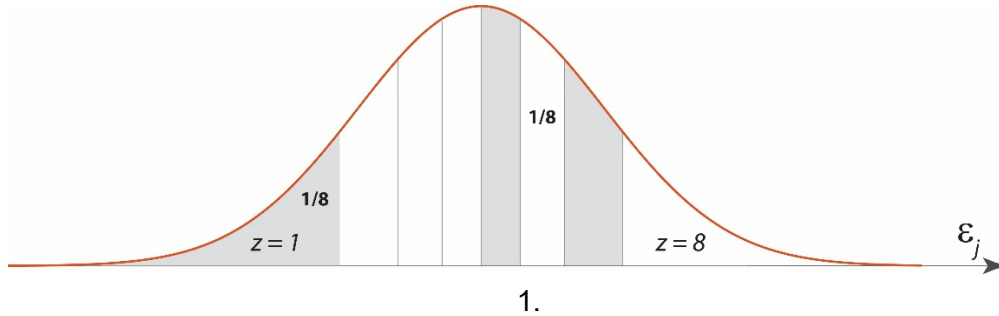


Figure 3-11: A plot of the fraction of the cumulative variance (eigenvalues) that contribute to the total variance (trace of the covariance matrix) for the North Appalachian Seismictectonic Zone. There is a total of 788 cells ($0.25^\circ \times 0.25^\circ$). Here, the sum of the first 375 eigenvalues out of 1576 represents 99% of the total variance.

Step 3: Generate eight realizations of ε to construct eight realizations of \tilde{X} using Equation 3-19. Each realization will represent an alternative recurrence map. The process is as follows:

1. Divide the sample space of each normally distributed ε_j into eight intervals, with each interval containing an equal probability of $1/8$. These intervals will have unequal lengths.



2. Within each interval z ($z = 1, 2, \dots, 8$), generate a random value ε_{jz} that follows the conditional distribution within the interval. The resulting values form a sequence with $\varepsilon_{j1} < \varepsilon_{j2} < \dots < \varepsilon_{j8}$ (Figure 3-12).

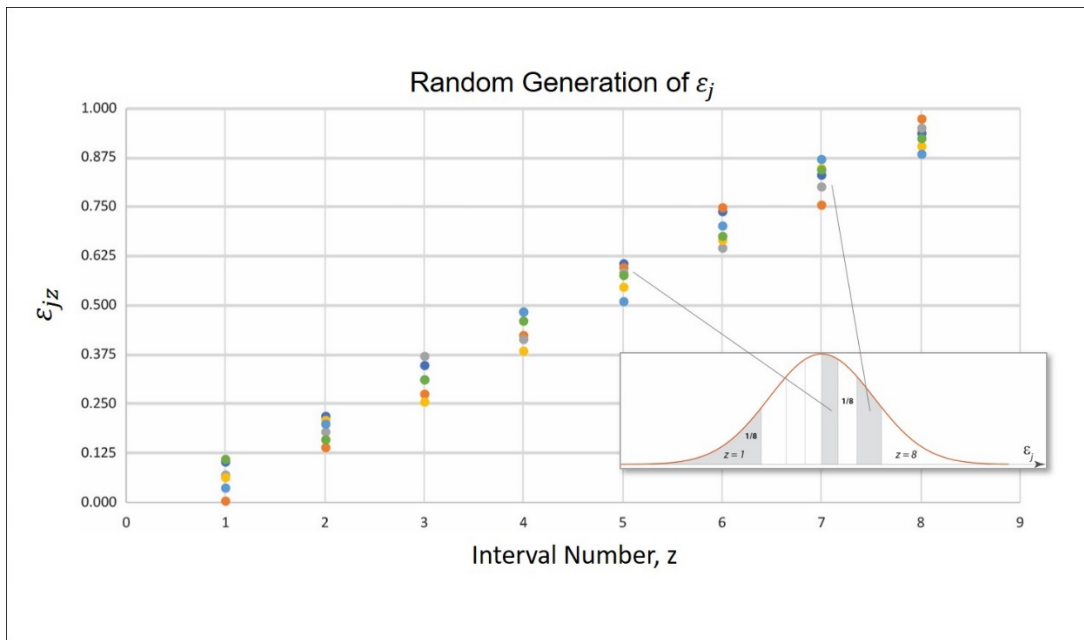


Figure 3-12: An example of random generation of 6 sets of ε_{jz} is shown here. Each set (color-coded) is sampled from 8 different segments of the normal distribution of ε_j with equal probability of $1/8$.

3. For each j , generate an independent random permutation of $\varepsilon_{j1}, \varepsilon_{j2}, \dots, \varepsilon_{j8}$. After this permutation, $\varepsilon_{j1}, \varepsilon_{j2}, \dots, \varepsilon_{j8}$ are no longer an increasing sequence. The result is eight random realizations of the random vector $\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k]$. Because the eight values for each j were drawn from the eight equal-probability intervals, the distribution of ε_j is well approximated by the eight realizations. Because of the random permutations, the

correlation coefficient between the eight realizations of any two ε_{j1} and ε_{j2} has an expected value of 0 and is likely to be low.

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{11} & \cdots & \varepsilon_{1k} \\ \vdots & \ddots & \vdots \\ \varepsilon_{81} & \cdots & \varepsilon_{8k} \end{pmatrix}$$

4. Generate eight alternative realizations of $\tilde{\mathbf{X}}$ using the eight permuted $\boldsymbol{\varepsilon}$ vectors and Equation 3-19.

$$\tilde{X}_j = \mathbf{m}_X + \sum_{k=1}^{2M+2} \varepsilon_{jk} \boldsymbol{\phi}_k$$

5. Convert each realization of $\tilde{\mathbf{X}}$ into values of $\ln v$ and β (first $2M$ elements of $\tilde{\mathbf{X}}$) for each cell in the source zone.
6. Use these two values to calculate the rate of earthquakes above magnitude 5 ($e^{\ln v} \times e^{-\beta(5-2.9)}$).

As an example, Figure 3-13 shows the eight alternative maps generated for NAP. The maps represent the uncertainty in recurrence parameters that results from the limited duration of the catalog. If the smoothing parameters are treated as uncertain and estimated objectively from the data, the eight alternative maps also include the uncertainty about the appropriate values of the smoothing parameters.

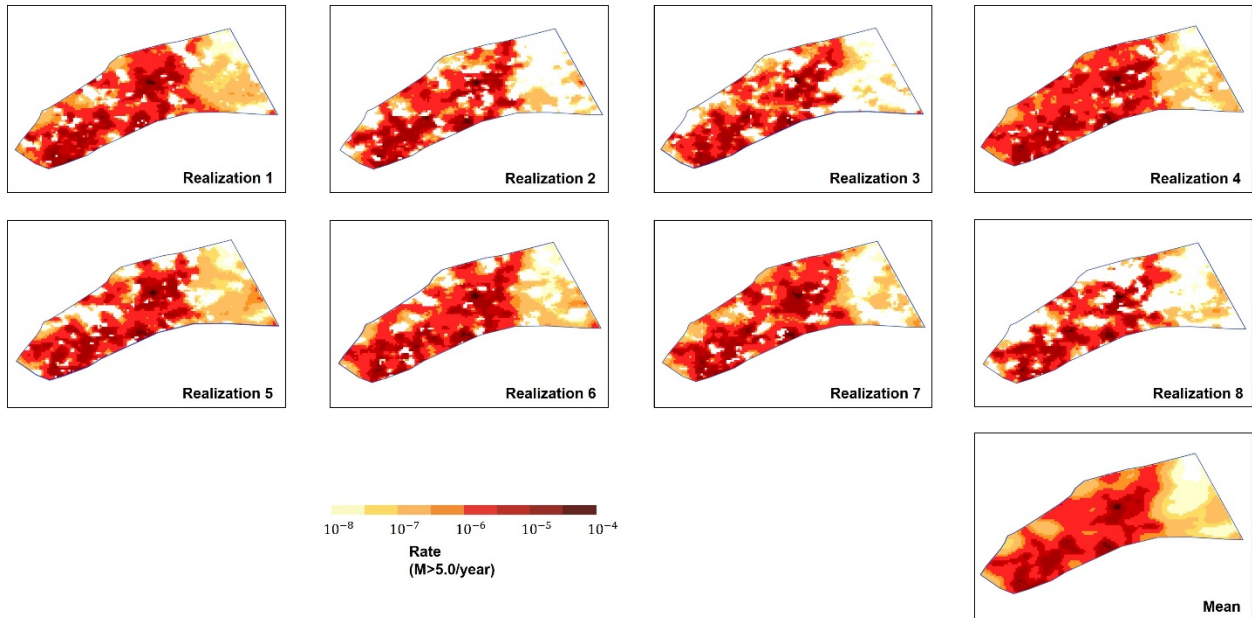


Figure 3-13: The mean and eight alternative maps of earthquake recurrence rate ($M_w \geq 5$) for the North Appalachian seismotectonic zone are shown here. The eight alternative maps capture the central tendency and statistical uncertainty in the recurrence parameters. The grid size is $0.1^\circ \times 0.1^\circ$.

As a check, a comparison of the mean rate map (average of all accepted MCMC realizations) and average of the eight alternative maps for the NAP seismotectonic is shown in Figure 3-14. As expected, they are very similar because the mean value of ε_k in Equation 3-19 is zero.

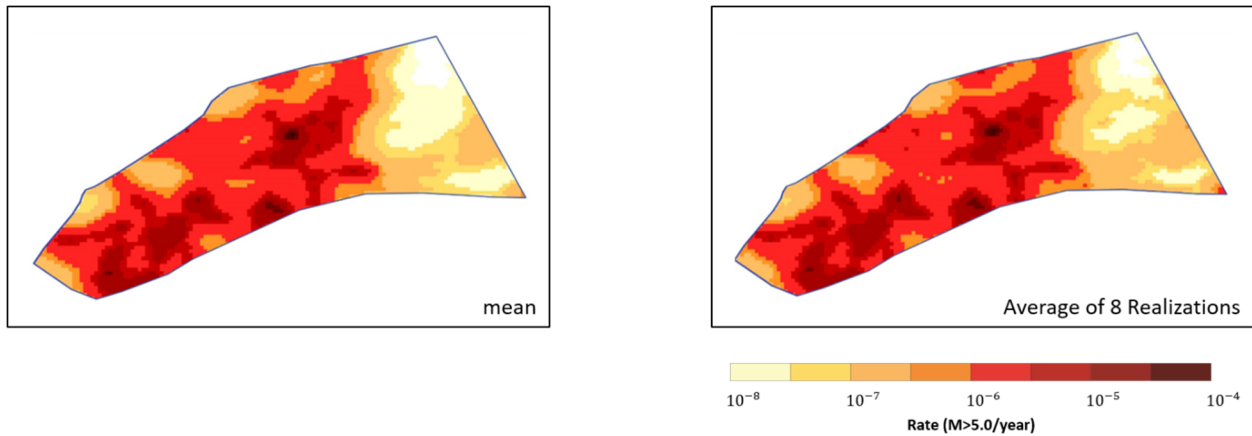


Figure 3-14: A comparison between the mean recurrence map ($M_w \geq 5$) and the average of the eight alternative maps for the NAP zone is shown here. The similarity between the two maps supports the notion that the alternative maps jointly represent the central tendency and statistical uncertainty in the recurrence parameters.

3.3.3 Application of the Model and Specification of Model Parameters

In addition to the algorithmic inputs described above, one needs to specify the weights for the various magnitude bins. Table 3-2 shows the five cases that were initially considered for the weights to the magnitude bins. These five alternative sets of weights span a wide range to investigate the effect of giving lower weights to the lower-magnitude bins. Case E was introduced later, as an intermediate case between Cases C and D, because Case D was considered too extreme.

Table 3-2: Alternative cases considered for the magnitude-dependent weights.

Case (wt)	Magnitude Bins					
	2.9 – 3.6	3.6 – 4.3	4.3 – 5.0	5.0 – 5.7	5.7 – 6.4	6.3 – 8.3
A (0.3)	1	1	1	1	1	1
B (0.3)	0.1	1	1	1	1	1
C	0	1	1	1	1	1
D	0	0.1	1	1	1	1
E (0.4)	0	0.3	1	1	1	1

Some of the reasons for giving lower weights to the lower magnitude bins are, for example, the magnitude-recurrence law may deviate from exponential (Figure 3-3), or the magnitude conversion models, or completeness model may be less reliable for lower magnitudes. In this regard, it is useful to note that only earthquakes with magnitudes greater than 5 are important for seismic hazard and risk for nuclear facilities. The only reason for considering lower magnitudes is that the $M_w > 4$ data alone are not sufficient for determining both the magnitude-recurrence law and the spatial distribution for earthquakes of engineering interest. Because the magnitude weights also affect the degree of smoothing, one also needs to consider the issue of spatial stationarity (i.e., whether the spatial pattern of past, small earthquakes is representative of the

spatial pattern of future, hazard-significant, earthquakes). Cases A through E were designed to cover a broad range of alternatives regarding magnitude weights.

Based on analyses in the CEUS-SSC project (NUREG-2115), only Cases A, B, and E were retained. The reasons for dropping Cases C and D were that case C was very similar to Case B and that Case D was considered too extreme, relying almost entirely on magnitudes 4.3 and greater, and leading to nearly spatially uniform seismicity. The three remaining cases (Cases A, B, and E) were examined and compared in terms of their fit to the magnitude-recurrence data and their degree of smoothing. Preference for the three cases was nearly equal, except that Case E was given slightly more weight because this case is more sensitive to data in the magnitude bins that that control seismic hazard. The resulting weights are 0.3, 0.3, and 0.4 for Cases A, B, and E, respectively.

Figure 3-15 shows the mean recurrence-rate maps ($M_W \geq 5$ /yr) for Cases A, B, and E calculated for NAP. The Case A map correlates closely with the seismic activity in the NAP zone from the 2018 USGS catalog (Figure 3-9). However, the recurrence rate maps become progressively more spatially uniform as smaller events are given lower weights (Cases B and E).

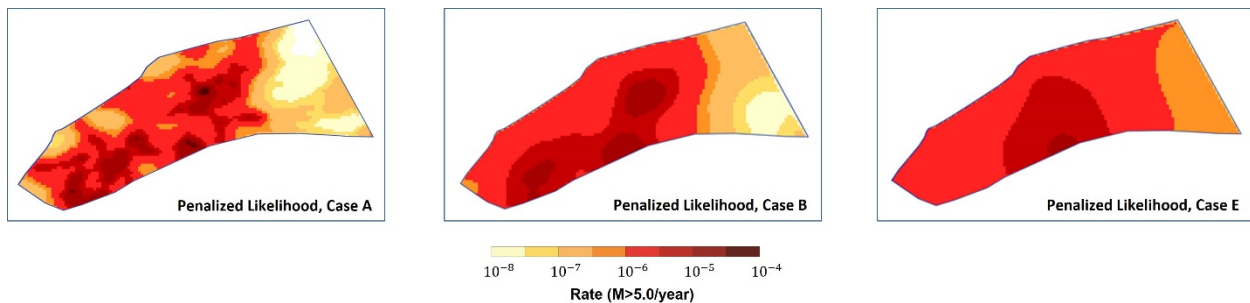


Figure 3-15: Maps of the mean earthquake recurrence rates ($M_W \geq 5$ /yr) for Cases A, B, and E estimated by using the Penalized-Likelihood smoothing method for North Appalachian (NAP) zone. The grid size in this case is $0.1^\circ \times 0.1^\circ$.

3.3.4 Effect of Cell Size on Estimated Recurrence Rate

To investigate the effect of grid size on the estimated earthquake recurrence rates, calculations were carried out for grid sizes $0.5^\circ \times 0.5^\circ$, $0.25^\circ \times 0.25^\circ$, and $0.1^\circ \times 0.1^\circ$.

Figure 3-16 shows the recurrence rates (weighted mean) for NAP seismotectonic zone for different grid sizes. As expected, spatial smoothing is more pronounced at larger grid sizes (lower resolution).

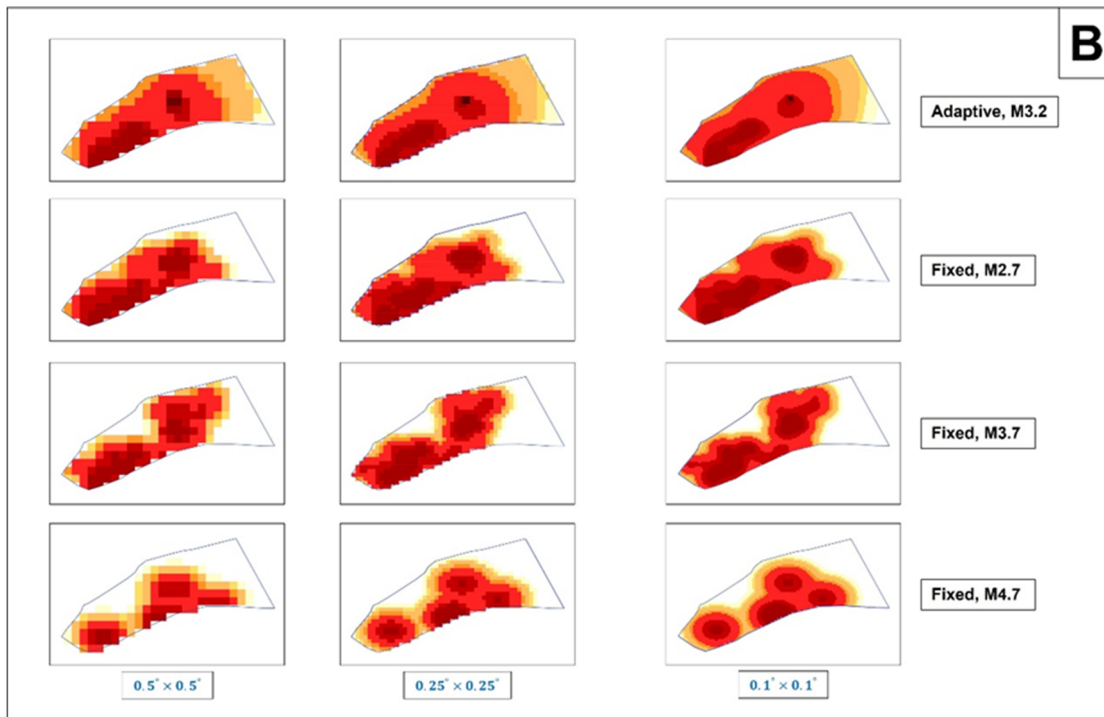
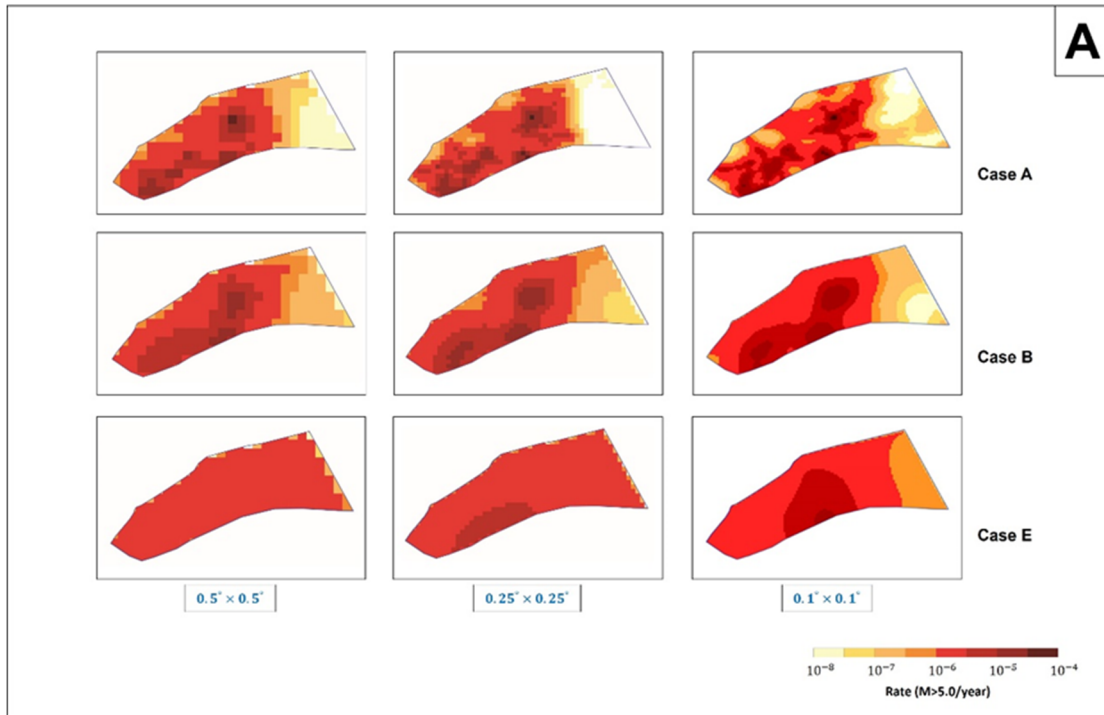


Figure 3-16: Mean recurrence rate maps of North Appalachian (NAP) zone calculated using three different grid sizes for (A) Penalized-Likelihood and (B) Kernel approaches. In general, the rate distributions remain the same with higher resolution at smaller grid size. However, in Case E (Panel A), the rate of smoothing tends to increase (more uniform) as the grid size increases.

4 SMOOTHING APPROACH COMPARISONS

In the 2018 National Seismic Hazard Modeling Project (NSHMP), the USGS used both fixed and adaptive kernel spatial smoothing approaches to estimate seismicity rates in the CEUS. In contrast, the CEUS-SSC project (NRC, 2012), a Senior Seismic Hazard Analysis Committee (SSHAC) project sponsored by the Department of Energy (DOE), Electric Power Research Institute (EPRI), and NRC, implements a penalized-likelihood smoothing approach that differs from that used by the USGS.

To highlight differences between the penalized-likelihood approach and the kernel methods, Section 5.3.2.3.3 of NUREG-2115 (NRC 2012) describes how kernel methods have a tendency to predict very low rates in regions with few or no recorded earthquakes because the commonly used Gaussian kernel function decays very rapidly at distances greater than twice the kernel width. This problem is discussed by Frankel et al. (1996), where the problem is avoided by introducing an alternative “floor” based on a constant weight. In contrast, the penalized-likelihood approach provides a natural floor for the activity rate. This can be inferred from Equation 3-3, where for $N=0$ (no recorded earthquake), the rate portion of the likelihood function reduces to an exponential distribution with an expectation (mean) value of $\nu_{ML} = \frac{1}{AT}$. This result also applies to one cell if there is no penalty function linking the cell’s parameters with its neighbors. This natural floor has the advantage that it does not require any external assumptions regarding the spatial extent that is used to determine the floor rate. Section 5.3.2.4 of NUREG-2115 (NRC 2012) provides additional discussions about differences between the Penalized-Likelihood approach and the kernel methods in estimating the recurrence rates and b -values, and reasons behind the CEUS-SSC excluding the kernel methods in the Probabilistic Seismic Hazard Analyses (PSHA) logic tree.

A study by Petersen et al. (2018) indicated that the different spatial smoothing approaches used by the USGS and the CEUS-SSC result in different seismic hazard estimates at specific sites in the CEUS, specially, those removed from RLME sources. Therefore, additional research was needed to evaluate differences in seismicity rates obtained from different smoothing approaches and the effects they might have on the evaluated hazard. As examples, Figure 4-3, Figure 4-2, and Figure 4-3 highlight the differences in the recurrence rate estimates using different smoothing approaches for the NAP and PEZ-W seismotectonic zones and the NMESE-N M_{max} zone, respectively. Figure 4-4, Figure 4-5, and Figure 4-6 compare $M_W \geq 5$ rates from NRC component models with USGS component models.

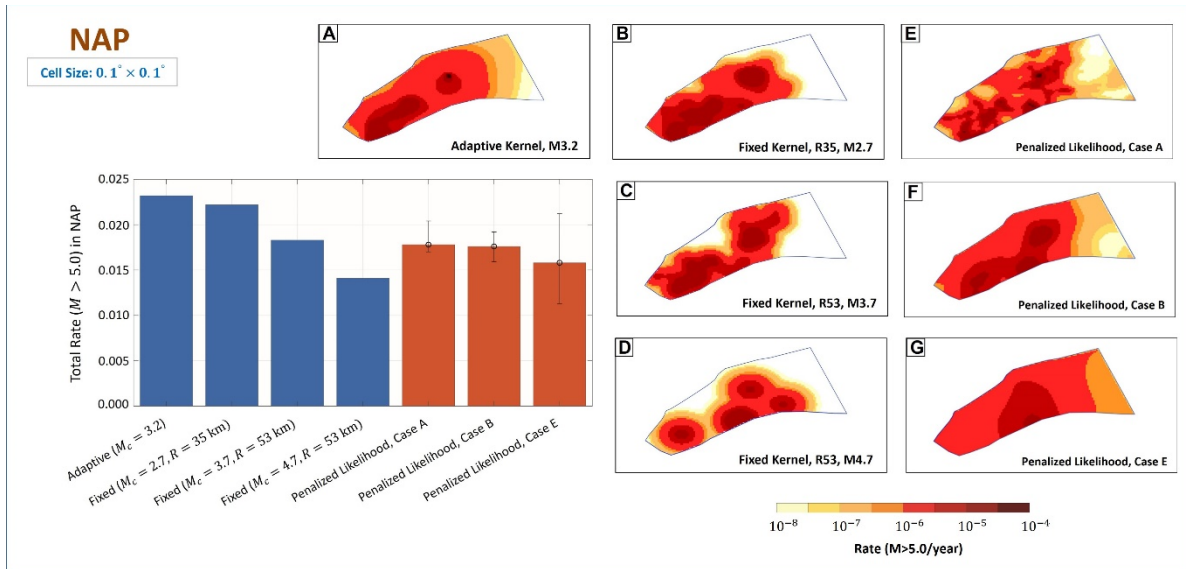


Figure 4-1: This figure shows differences in estimates of the recurrence rates for $M_W \geq 5.0$ earthquakes in North Appalachian (NAP) using the adaptive kernel (A), the fixed kernel (B-D), and the penalized-likelihood (E-G) smoothing approaches. The chart shows the total zone rates ($M_W \geq 5.0$) for all approaches. The error bars show the range covered by the eight realizations in the penalized-likelihood method. In all cases, the 2018 U.S. Geological Survey (USGS) earthquake catalog is used as the input.

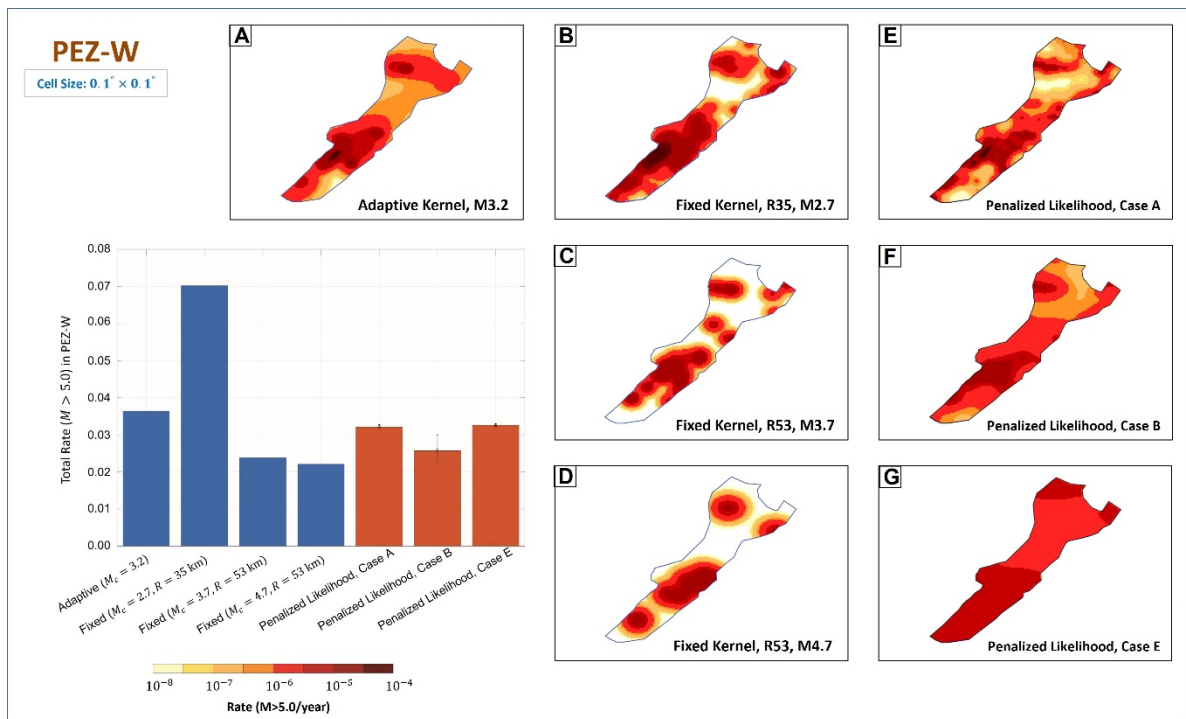


Figure 4-2: This figure shows differences in estimates of the recurrence rates for $M_W \geq 5.0$ earthquakes in Paleozoic Extended Crust-Wide (PEZ-W) using the same smoothing approaches and inputs for each subpanel A-G as described for Figure 4-1.

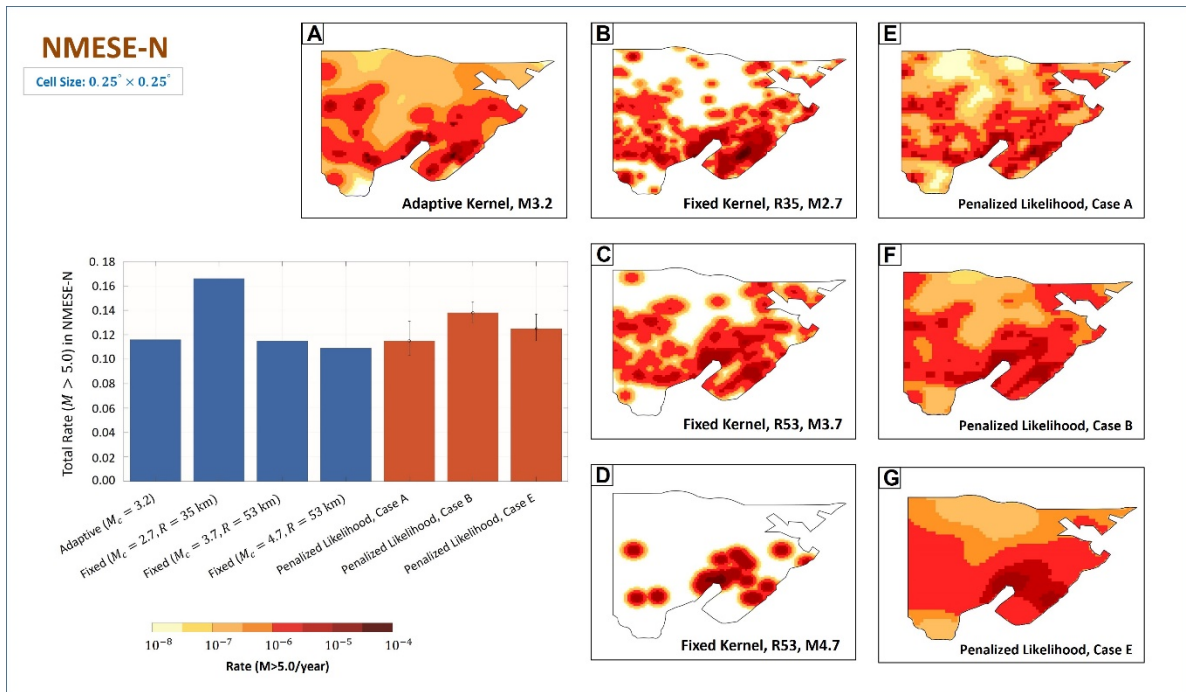


Figure 4-3: This figure shows differences in estimates of the recurrence rates for $M_W \geq 5.0$ earthquakes in Non-Mesozoic-and-Younger Extension-Narrow (NMESE-N) using the same smoothing approaches and inputs for each subpanel A-G as described for Figure 4-1.

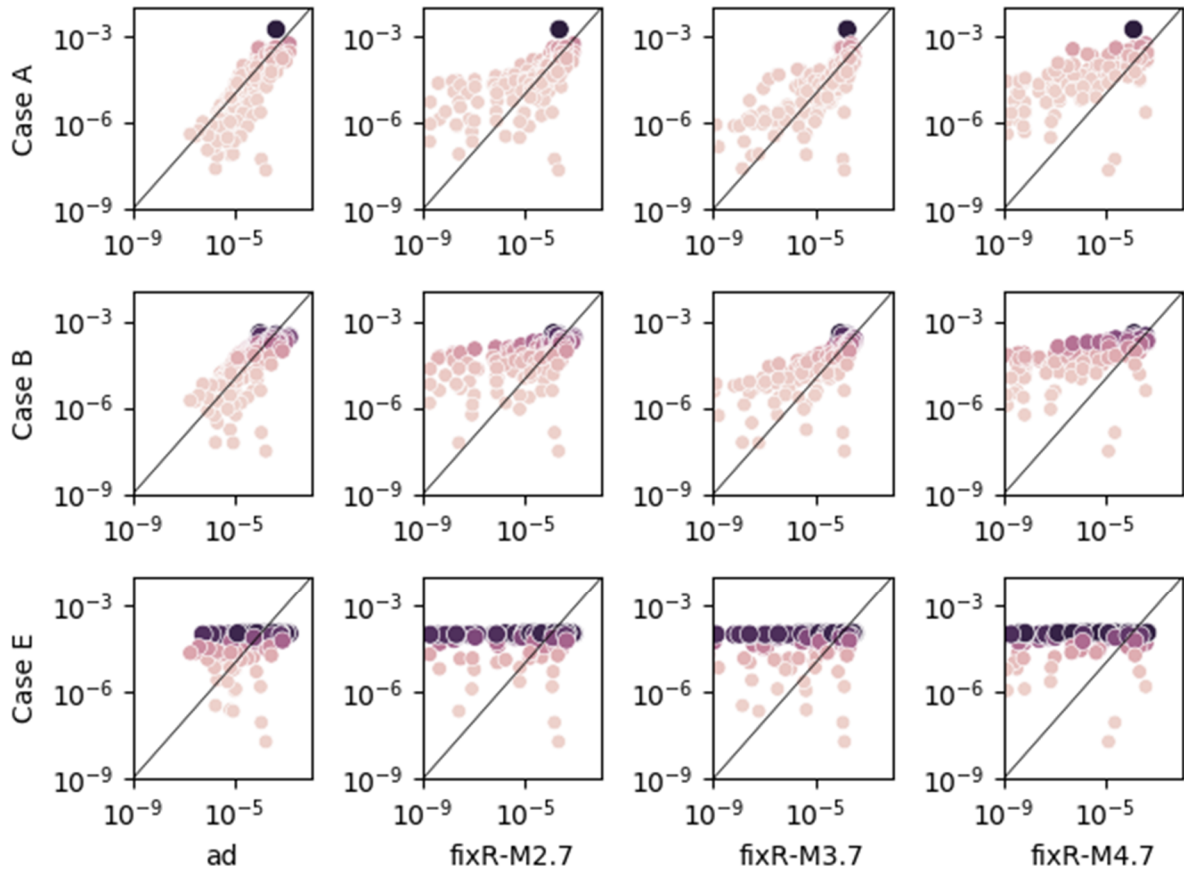


Figure 4-4: This figure depicts estimates of the recurrence rates for $M_W \geq 5.0$ earthquakes in North Appalachian (NAP) between Nuclear Regulatory Commission (NRC) and U.S. Geological Survey (USGS) rate models. NRC rate models include Case A, Case B, and Case E. USGS rate models include the adaptive smoothing (ad) and fixed smoothing with $M_c = 2.7$ (fixR-M2.7), $M_c = 3.7$ (fixR-M3.7), and $M_c = 4.7$ (fixR-M4.7).

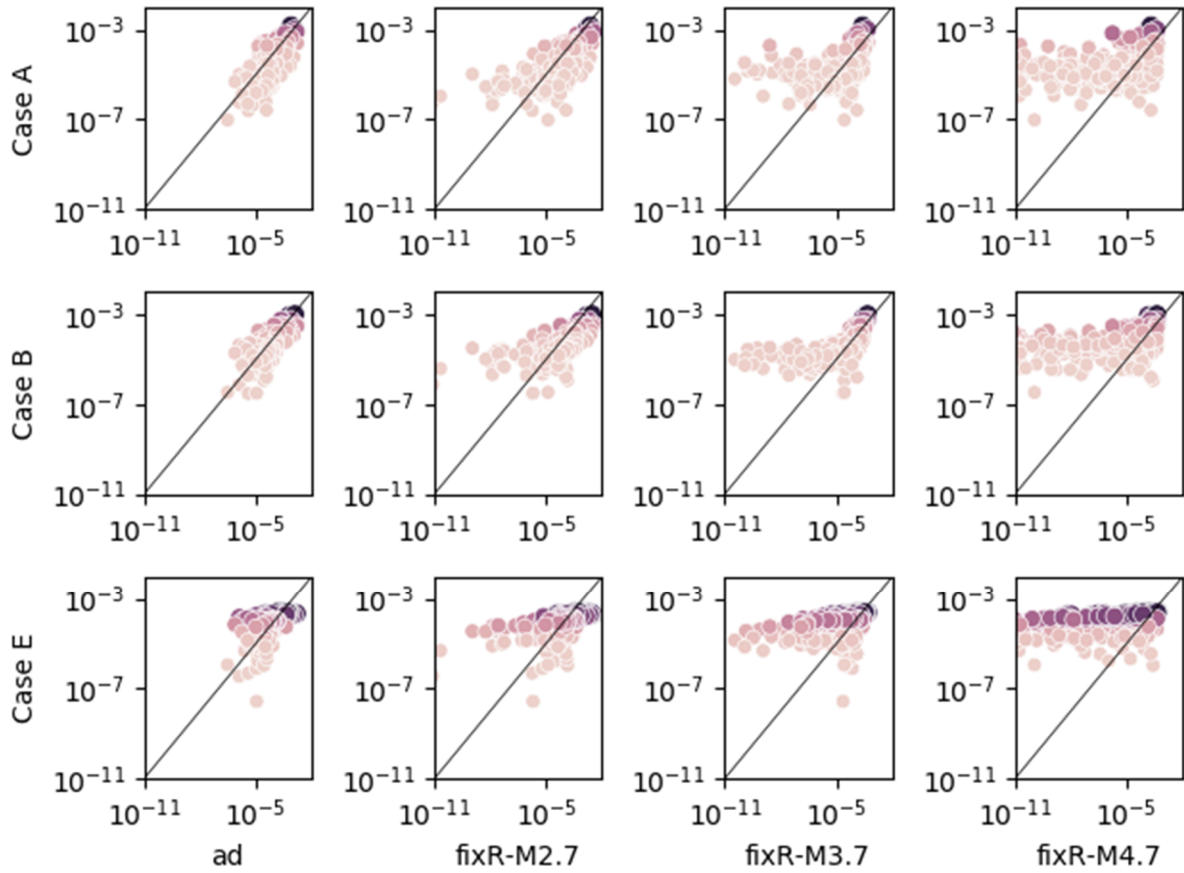


Figure 4-5: This figure depicts estimates of the recurrence rates for $M_W \geq 5.0$ earthquakes in Paleozoic Extended Crust-Wide (PEZ-W) between Nuclear Regulatory Commission (NRC) and U.S. Geological Survey (USGS) rate models. All other model details are the same as in Figure 4-4.

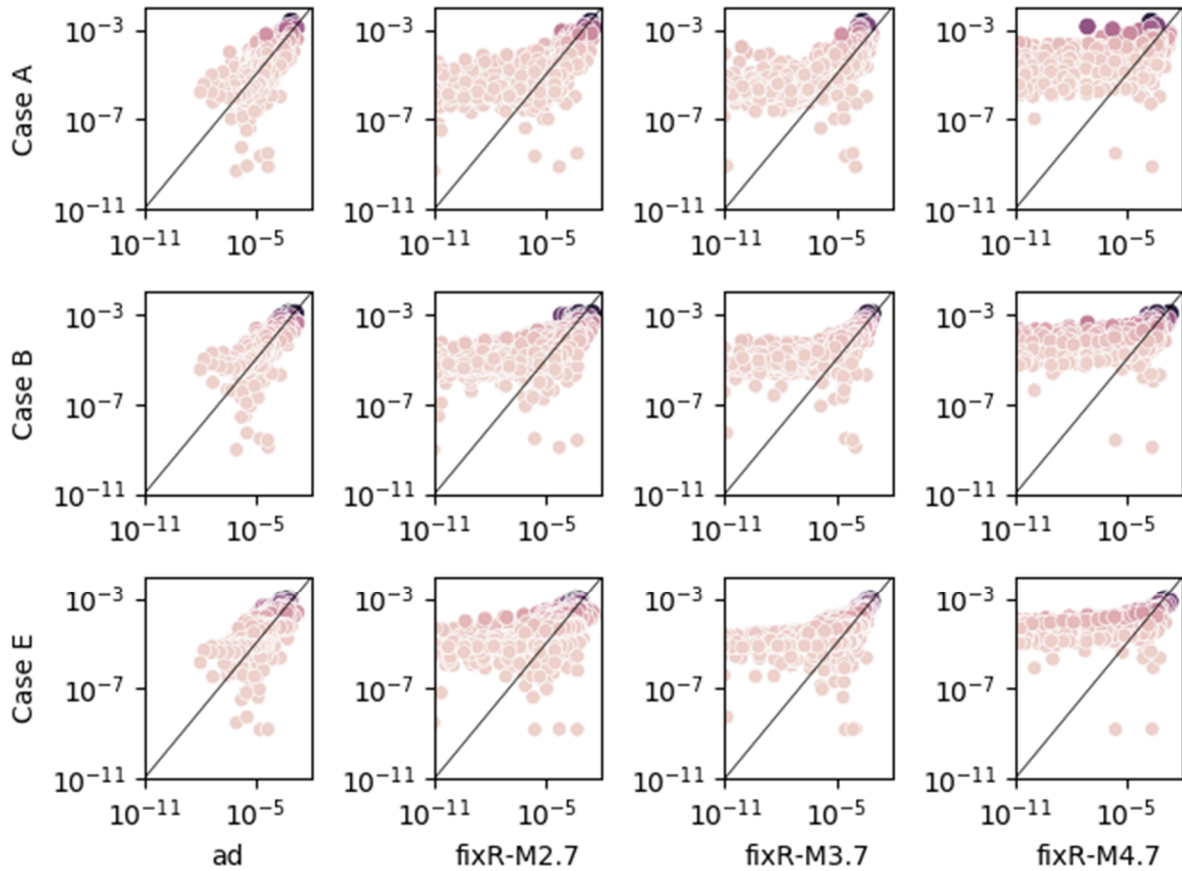


Figure 4-6: This figure depicts estimates of the recurrence rates for $M_W \geq 5.0$ earthquakes in Non-Mesozoic-and-Younger Extension-Narrow (NMESE-N) between Nuclear Regulatory Commission (NRC) and U.S. Geological Survey (USGS) rate models. All other model details are the same as in Figure 4-4.

4.1 Catalog and Catalog Completeness

We use a common earthquake catalog for the comparison of spatial smoothing methods. The catalog was developed using the methods adapted for the 2014 and 2018 National Seismic Hazard Model (NSHM) updates (Mueller, 2019; Petersen et al., 2015, 2020). The earthquake catalog is available at <https://doi.org/10.5066/P95SNP2J> (Mueller, 2023). The approach comprises combining source catalogs, converting earthquake magnitudes to moment magnitude, and eliminating duplicate and human-induced events to produce a uniform moment magnitude catalog for recurrence estimates in the CEUS. Earthquake magnitude conversions use the regionalization and magnitude-type-dependence specified by CEUS-SSC. We use the Gardner and Knopoff (1974) algorithm to decluster the catalog so that the seismicity-rate estimates correspond to independent events. The catalog builds on the earthquake catalog used in the 2018 NSHM (Petersen et al., 2020) and covers the period 1568–2018. There are 4532 events with $M_W \geq 2.7$, 1906 events with $M_W \geq 3.2$, 768 events with $M_W \geq 3.7$, and 108 events with $M_W \geq 4.7$.

Beside implementing different spatial smoothing methods for estimating earthquake recurrence rates, the CEUS-SSC and USGS subdivide the CEUS into geographic subregions (14 zones by the NRC, and 7 zones by the USGS), within which the earthquake catalog for a specific date and

magnitude range (or a minimum magnitude) is complete. The following sections describe how subregions and completeness are defined by the CEUS-SSC and the USGS.

The assessment of earthquake catalog completeness is needed to prevent underestimation of earthquake recurrence rates that are important in seismic hazard assessment.

The process of testing the catalog completeness provides estimates of the time (calendar year) when all the earthquakes in a given magnitude range (or above a minimum magnitude, M_c), and in a seismotectonic zone have been accounted for. The length of catalog completeness is typically a function of magnitude, with larger magnitudes having longer completeness periods. The data identified in this way would be used to assess recurrence parameters using a procedure such as Weichert's (1980) maximum likelihood formulation for binned magnitude data. The following sections provide a brief description of the USGS and CEUS-SSC approaches in estimating the starting year of a completeness level and the duration of the completeness level for a given magnitude bin (or a magnitude threshold, M_c).

4.1.1 USGS Approach

In developing earthquake rate models, the USGS specifies time periods and minimum magnitude thresholds that guide the selection of earthquakes from the catalog. The magnitude threshold used is commonly the minimum magnitude of completeness (M_c), which is the minimum magnitude for which the catalog reports all earthquakes. By specifying that the ending year of all completeness levels is the ending year of the catalog (e.g., the present) and assuming that M_c does not diminish with time, each completeness level can therefore be characterized by a starting year and an M_c value. The starting year of a completeness level is referred to as a "completeness year" and the duration of the completeness level as the "completeness duration." Once the completeness years are specified, the corresponding M_c values are calculated using the methods discussed in Moschetti et al. (2015). This approach differs from the one used for the CEUS component of the U.S. NSHM (Petersen et al., 2014), which specified regional M_c values and then computed corresponding completeness years.

The USGS divides the CEUS into six geographic zones of completeness within which M_c is assumed to be uniform. These zones are labeled as ZC1–ZC6. The region outside the six zones is labeled as ZC_{out} (Figure 4-7). These subregions are assumed to contain sufficient numbers of earthquakes for estimating M_c . Completeness zones were taken from Petersen et al. (2014); these zones were developed with methods originally described in a CEUS seismic-hazard analysis by EPRI (1988) and modified in EPRI/DOE/NRC (2012). Earthquakes occurring outside of the six defined zones (ZC1–ZC6) were analyzed separately (ZC_{out}). The two tables in Figure 4-7 list the "completeness year" in each zone for both the fixed- and adaptive-kernel methods.

Smoothing type (i.e., fixed and adaptive) and smoothing-kernel bandwidth are correlated with the completeness levels in the USGS model. Minimum magnitudes of completeness of $M_c = 2.7, 3.7, 4.7$ are used in fixed smoothing models with smoothing-kernel bandwidths of 35, 53, and 53 km, respectively. Adaptive smoothing uses $M_c = 3.2$ only.

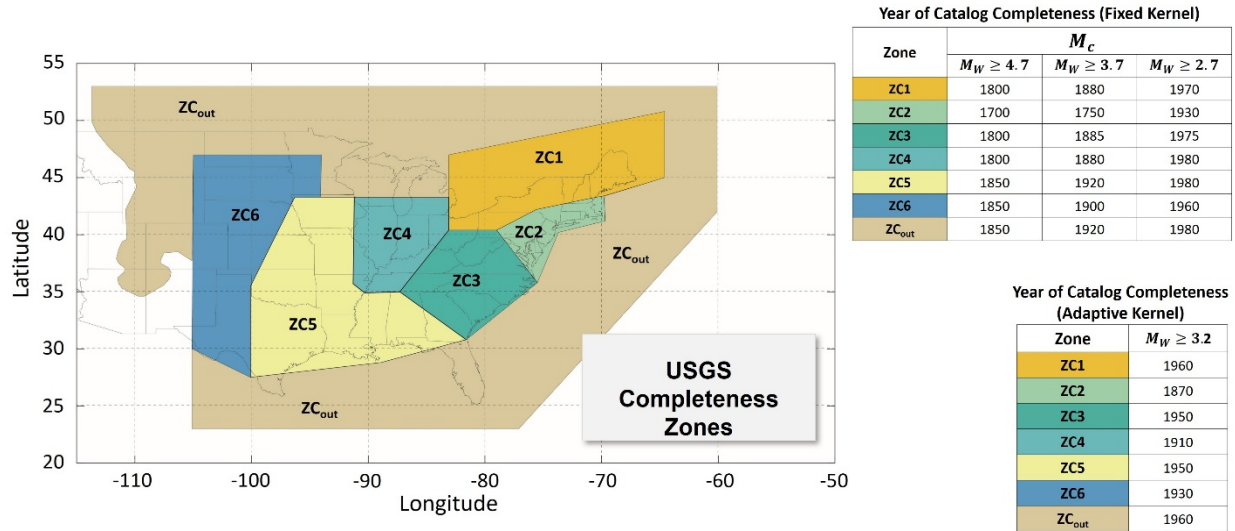


Figure 4-7: The seven zones of completeness (ZC1–ZC6, ZC_{out}) used in the U.S. Geological Survey (USGS) smoothing approach to generate earthquake recurrence rate maps. The tables list M_c and the years of catalog completeness for both fixed and adaptive kernel methods in each zone.

4.1.2 CEUS-SSC Approach

To determine the catalog completeness, CEUS-SSC uses the approach first proposed by Stepp (1972). This approach evaluates the catalog completeness for specific magnitude ranges by starting at the present (t_p) and moving back in time and counting the total number of earthquakes in the catalog in each magnitude interval. At each point in time ($t_i(m)$) when an earthquake in the specified magnitude interval (m) occurred, the annual rate of earthquakes in the magnitude interval ($v_i(m)$) is computed by dividing the sum of the number of earthquakes from that point in time to the end of the catalog ($n_i(m)$) by the length in time from that point to the end of the catalog ($t_p - t_i(m)$). Assuming that the rate of earthquakes is constant in time, plotting these values versus date for the complete portion of the catalog will show an approximately horizontal line (Figure 4-8). As one moves farther back in time, eventually the plotted line will start to trend downward, indicating that not all earthquakes are being reported (again assuming stationarity in time of the true rate). The point at which this downward trend begins indicates the beginning of the complete period of catalog reporting for the specific magnitude interval, $t_c(m)$.

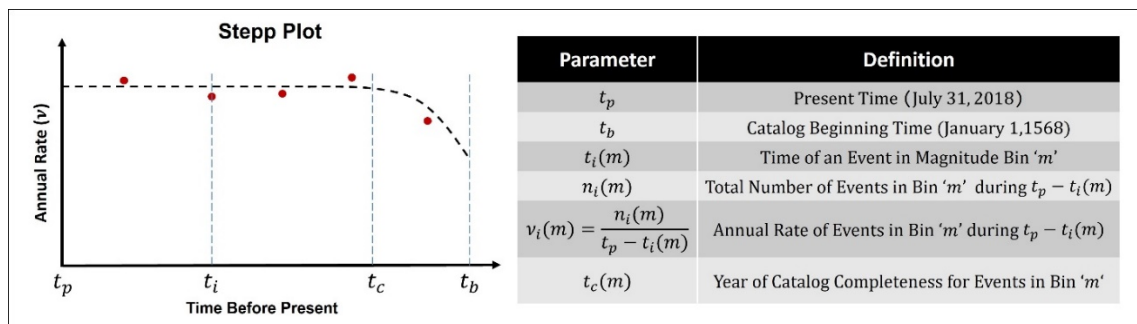


Figure 4-8: A schematic showing how catalog completeness is determined through Stepp (1972) approach. Dates for t_p and t_b are based on the 2018 U.S. Geological Survey (USGS) catalog.

In Stepp plots, using only the complete portion of the catalog (from t_c to present time, t_p) has the potential of not including earthquake data from the incomplete periods of the catalog. To account for the catalog data in the partially complete periods, the EPRI-SOG Project (EPRI, 1988) defined a parameter called the probability of detection, P_D (Equation 3-6), that represents the probability that an earthquake in any point in time would be recorded and appear in the seismic record. This approach was used by CEUS-SSC to calculate the equivalent period of completeness, T_E , in the CEUS.

In the penalized-likelihood approach, catalog completeness is determined by dividing the CEUS into 14 geographic regions (Figure 4-9, Frame B). EPRI (1988), through analysis of the history of population growth and earthquake recording, defined 13 completeness regions covering most of the CEUS. An additional Completeness Region 14 was added to cover the Gulf of Mexico, as offshore earthquakes in that area are important to the assessment of seismic hazards along the Gulf Coast. These regions represent portions of the CEUS where catalog completeness as a function of time and magnitude is assessed to be sufficiently similar such that it can be treated as the same. As an example, Table 4-1 shows the probability of detection and the equivalent periods of completeness for Region 12. The probability of detection values listed in Tables 3.5-1, 3.5-2, and 3.5-3 of NUREG-2115 (NRC 2012) were used to calculate T_E . As expected, the equivalent period of completeness is longer for larger magnitude bins.

Table 4-1: Probability of detection and equivalent periods of completeness for Region 12 (Figure 4-9) for Case B (extracted from Table 3.5-2, NUREG 2115; NRC 2012).

Magnitude Bin	Probability of Detection for Time Period (P_D)							Equivalent Period of Completeness (T_E , yr)	Beginning Year of Usable Period
	1625–1780	1780–1860	1860–1910	1910–1950	1950–1975	1975–1995	1995–2018		
2.9–3.6	0	0.04	0.27	0.293	0.506	1	1	84.1	1780
3.6–4.3	0	0.121	0.415	0.415	1	1	1	115.0	1780
4.3–5.0	0	0.619	1	1	1	1	1	207.5	1625
5.0–5.7	0	1	1	1	1	1	1	238.0	1625
5.7–6.4	0	1	1	1	1	1	1	238.0	1625
6.4–8.3	0	1	1	1	1	1	1	238.0	1625

Figure 4-9 shows a plot for earthquake recurrence rate as a function of time for catalog completeness (A) in the CEUS-SSC Region 12 (B). The table on the lower right shows the beginning year of *usable* period and the *equivalent* period of completeness for Case B in Region 12 calculated from the USGS 2018 earthquake catalog.

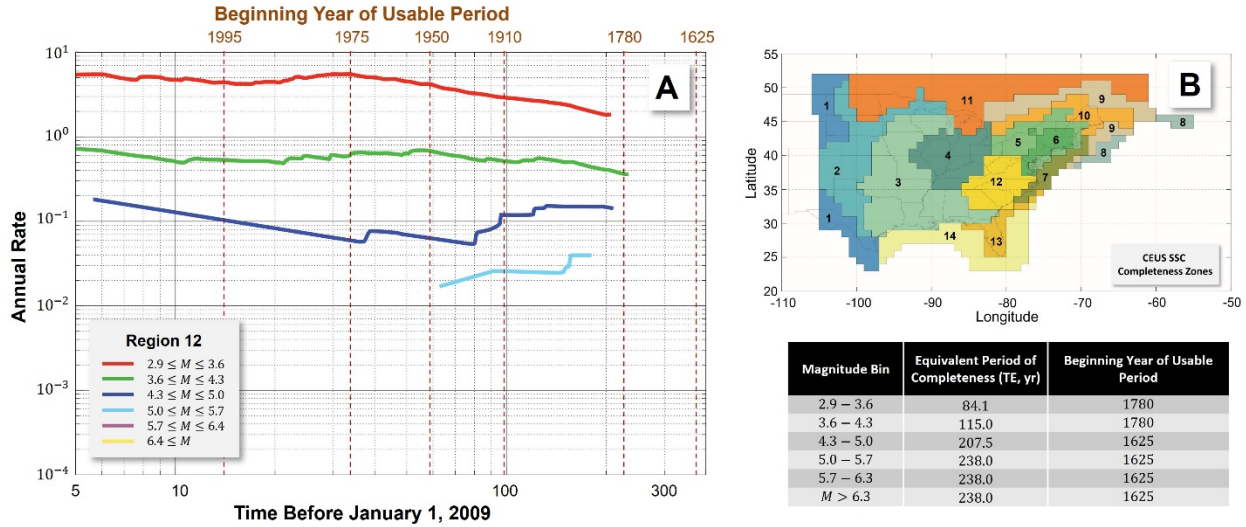


Figure 4-9: The “Stepp” plot of earthquake recurrence rate is shown in (A) as a function of time for catalog completeness in Region 12 (Figure 3.5-4, NUREG-2115; NRC 2012). The 14 regions of completeness designated for the Central and Eastern United States Seismic Source Characterization (CEUS-SSC) are shown in (B). The table shows the “beginning year of usable data” and the updated “equivalent time of completeness” in Region 12 using the U.S. Geological Survey (USGS) 2018 earthquake catalog (Case B, Table 3-2).

In order to check the sensitivity of estimated earthquake recurrence rate to different catalog completeness criteria, the mean recurrence-rate maps ($M_w \geq 5$) for NAP seismotectonic zone (Case A, defined in Table 3-2) were calculated using the penalized-likelihood method with the CEUS-SSC and the USGS zones of completeness. Figure 4-10 shows the three maps and the chart of total recurrence rates ($M_w \geq 5$) corresponding to the three maps. Although there are some differences, the distributed seismicity looks very similar in all three maps.

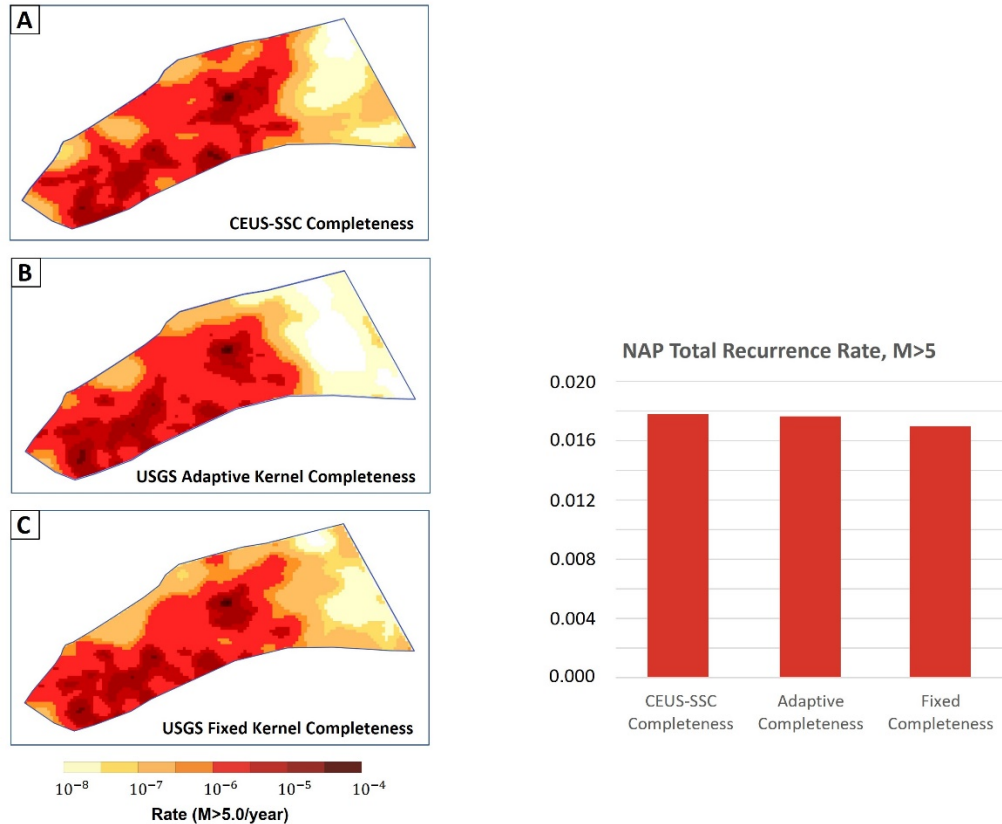


Figure 4-10: Maps in the left column show the mean earthquake recurrence rates ($M_w \geq 5$) in the North Appalachian (NAP) seismotectonic region calculated using the Penalized-Likelihood approach for Case A (see Table 3-2). Recurrence rates in (A-C) are calculated using the seismic zones and completeness periods of the Central and Eastern United States Seismic Source Characterization (CEUS-SSC), U.S. Geological Survey (USGS) Adaptive Kernel (AK), and USGS Fixed Kernel (FK), respectively. Although there are some differences, the distributed seismicity looks very similar in all three maps. This is supported by the chart on the right, which shows almost identical total recurrence rates in NAP for all three cases.

4.2 Comparison of Recurrence Parameters to Catalog

The Gutenberg-Richter formula (Equation 3-1) represents a continuous distribution of earthquake magnitudes. However, event magnitudes can be determined within a finite accuracy. In addition, Equation 3-1 assumes that the exponential nature applies for both small and large magnitude earthquakes. But, as shown in Figure 3-3, in practice the exponential assumption may not apply at small scales (i.e., over differences of a few tenths of a magnitude unit, or less). Therefore, to account for the discrete nature of the earthquake catalog, and to reduce sensitivity to small-scale deviations from the exponential model, the likelihood function is reformulated in terms of the earthquake counts in discrete magnitude bins (Equation 3-5). The likelihood function is further modified to account for maximum magnitude M_{\max} (Equation 3-8). Because M_{\max} could vary in different seismotectonic zones, it would need to be independently estimated based on geophysical considerations such as maximum fault lengths, regional stress drop, and earthquake history (Weichert, 1980).

Weichert (1980) developed a maximum likelihood approach for estimating seismicity rates and characterizing uncertainty in the rate. Rates for an entire seismotectonic zone were computed using this approach as one way to assess results from the penalized-likelihood and kernel methods. In this report, the CEUS-SSC estimated recurrence parameters (ν_{ML} and β_{ML}) are compared with those determined from the catalog using Weichert's approach for a check on consistency.

4.2.1 Estimates of Earthquake Recurrence Parameters (Weichert, 1980)

In seismic hazard studies, the recurrence rate parameters (ν and β) are estimated from the earthquake catalogs using the maximum likelihood of the Gutenberg-Richter relation. Using Equation 3-3, the maximum likelihood estimates of ν_{ML} and β_{ML} are given by:

$$\nu_{ML} = \frac{N}{AT} \quad 4-1$$

and

$$\frac{1}{\beta_{ML}} = \frac{1}{N} \sum_i m_i - m_0. \quad 4-2$$

As shown in Figure 4-11, the maximum likelihood estimates of b -value have sharper peaks (i.e., lower coefficient of variation) when N is large, indicating that only a narrow range of b -values (or ν) are consistent with the catalog.

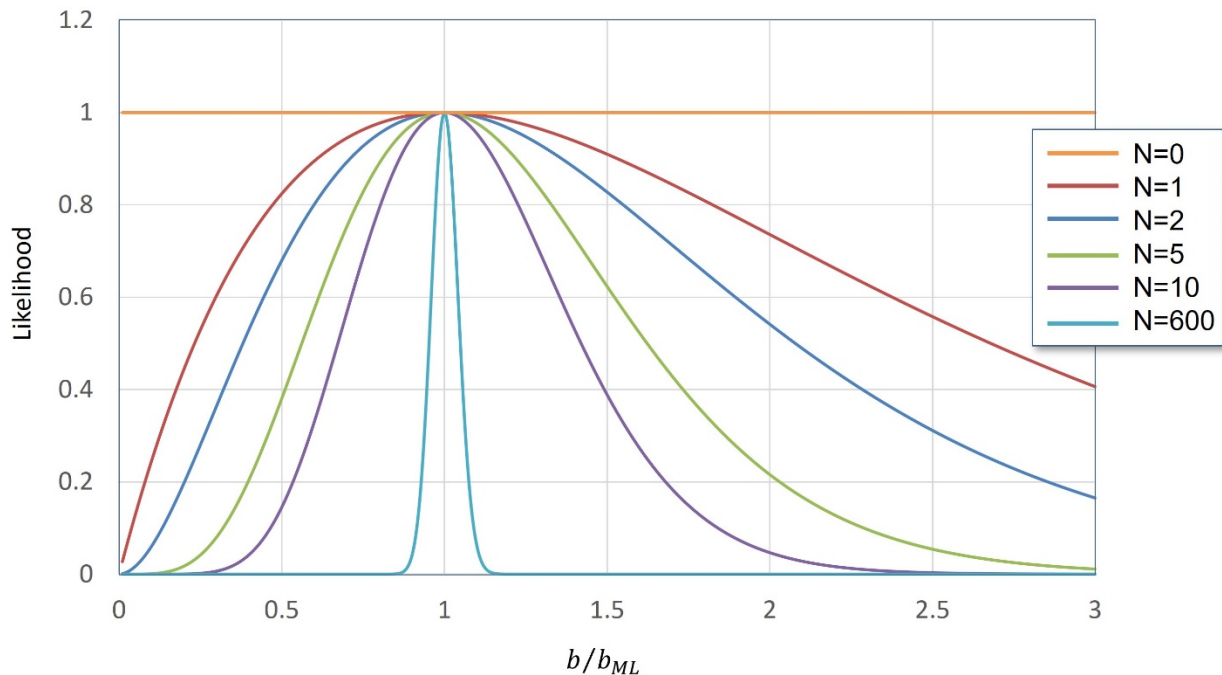


Figure 4-11: Likelihood function for b -value of an exponential magnitude distribution, for multiple values of the earthquake count N . The value of b is normalized by the maximum-likelihood estimate b_{ML} . (Figure 5.3.2-2, NUREG-2115; NRC 2012)

The above estimates are derived for continuous magnitude distributions without an upper bound for earthquake magnitude. However, earthquake magnitudes in the catalog are essentially discrete, have a possible regional upper bound (M_{\max}), and depending on their size, have different periods of completeness.

To determine estimates of β_{ML} from the earthquake catalog, the approach of Weichert (1980) is used here, which makes the following assumptions:

- a) group earthquakes in magnitude range $m_i \pm \delta$ in bin 'i',
- b) impose a regional maximum magnitude, M_{\max}
- c) assign unequal observational periods (period of completeness), t_c^i , using the available techniques (e.g., Stepp, 1972).

Assuming a truncated recurrence density, the probability $p(m)$ of observing an earthquake with magnitude between m and $m + dm$ is given by (Weichert, 1980):

$$p(m) dm \propto \begin{cases} e^{-\beta m} dm & m_0 \leq m \leq m_{\max} \\ 0 & \text{otherwise} \end{cases}$$

Then, the likelihood function becomes

$$L(\beta | n_i, m_i, t_c^i) = \frac{N!}{\prod n_i!} \prod p_i^{n_i}, \quad 4-3$$

where

$$p_i = \frac{t_c^i e^{-\beta m_i}}{\sum_j t_c^j e^{-\beta m_j}}.$$

Equating partial derivative of $\ln L$ with respect to β to zero provides the extremum condition (Equation 6 in Weichert, 1980):

$$\sum_i n_i m_i - N \frac{\sum_j t_c^j m_j e^{-\beta m_j}}{\sum_j t_c^j e^{-\beta m_j}} = 0. \quad 4-4$$

Using an iterative method, the equation can be solved for β_{ML} .

Figure 4-12 shows comparisons between the penalized-likelihood predicted b -values for the eight recurrence rate realizations in the NAP and PEZ-W zones using Cases A, B, and E magnitude weights with those obtained from the Weichert (1980) approach. In these calculations, for the sake of consistency, the equivalent period of completeness, T_E , is used in place of Weichert's unequal observational period, t_c^i . The error bars represent the 16%–84% uncertainty associated with the earthquake data (Weichert, 1980). For comparison, b -values ($b = 1$) for the USGS adaptive- and fixed-kernel approaches also are shown. These comparisons indicate a good agreement between predicted and observed rates for magnitudes of 5 and greater.

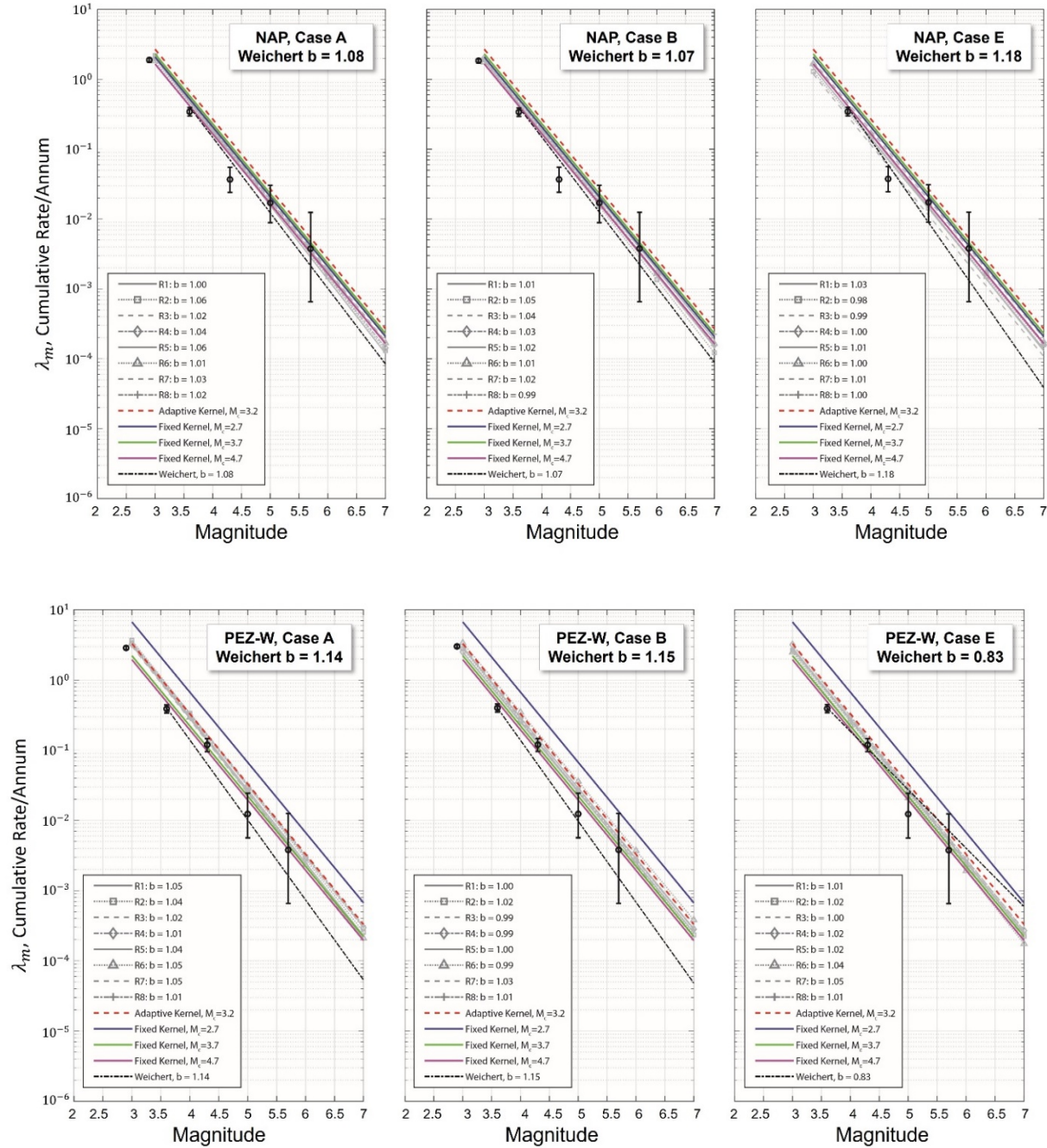


Figure 4-12: Penalized-likelihood predicted b -values for the eight recurrence rate realizations in the North Appalachian (NAP) and Paleozoic Extended Crust-Wide (PEZ-W) zones using Cases A, B, and E magnitude weights are compared with those obtained from the Weichert (1980) approach and using the equivalent period of completeness, T_E , in place of Weichert's unequal observational period, t_c^j in Equation 4-4. The error bars represent the 16%–84% uncertainty associated with the earthquake data (Weichert, 1980). For comparison, the unity b -values the USGS adaptive- and fixed- kernel approaches also are shown. (Note that the 'Weichert b ' line is drawn to show the slope only, not the actual magnitude distribution.)

4.3 Hazard Curves

To compare hazards calculated from different smoothing approaches, in each of the three seismotectonic zones, three arbitrary sites were selected (Table 4-2). Then, using the seismicity rate within a 400-km radius about each of the sites, hazard curves were calculated for 24 penalized-likelihood rate realizations and four adaptive and fixed-kernel rate realizations. The hazard curves were calculated for 1 Hz and 10 Hz spectral accelerations using the logic trees for the three CEUS-SSC source zones and for the ground motion model (GMM) for Eastern North America called Next Generation Attenuation Relationships for Central & Eastern North America (NGA-East) (Goulet et al., 2018). Based on the implementation of the logic trees for the source zones and GMM, there are 6,120 hazard curves for the implementation of the penalized-likelihood rate realizations and 255 hazard curves for the four adaptive and fixed-kernel rate realizations. The substantially larger number of hazard curves for the implementation of the penalized-likelihood rate realizations is due to the 24 alternative sets of recurrence rate parameters captured in the CEUS-SSC logic tree for each of the gridded cells in the three source zones. Figure 4-13, Figure 4-14, and Figure 4-15 show the hazard curves computed at 10 Hz spectral acceleration for select sites in NAP, PEZ-W, and NMESE-N, respectively. The results generally indicate the penalized-likelihood hazard curves envelop those obtained from the adaptive and fixed kernel seismicity rates. Mean hazard curves are also shown in these figures as the thick solid lines.

Table 4-2: Locations of sites where hazard curves were calculated.

Seismotectonic Zone	Location of Hazard Calculation	
	Latitude	Longitude
NAP	46.0°	-68.9°
PEZ-W	35.5°	-83.2°
NMESE-N	40.0°	-96.0°

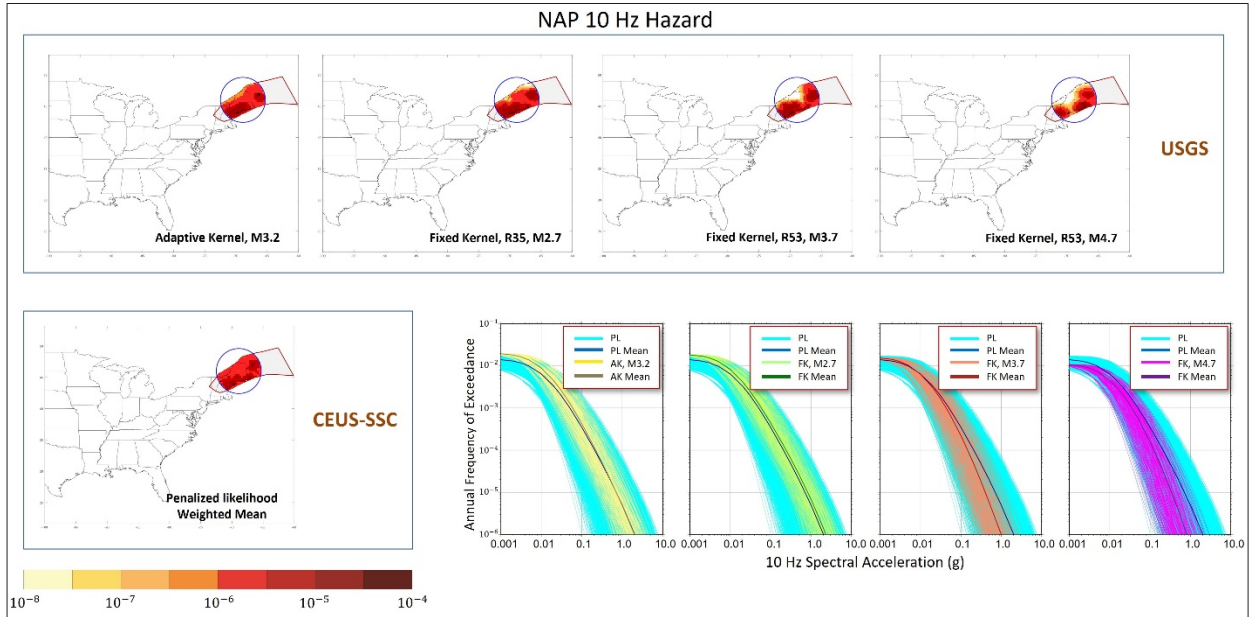


Figure 4-13: Hazard curves computed at 10 Hz spectral acceleration at a select site in the North Appalachian (NAP) seismotectonic zone (see Table 4-2 for coordinates) using Adaptive Kernel (AK), Fixed Kernel (FK), and Penalized-Likelihood (PL) methods. Maps show results computed with different approaches.

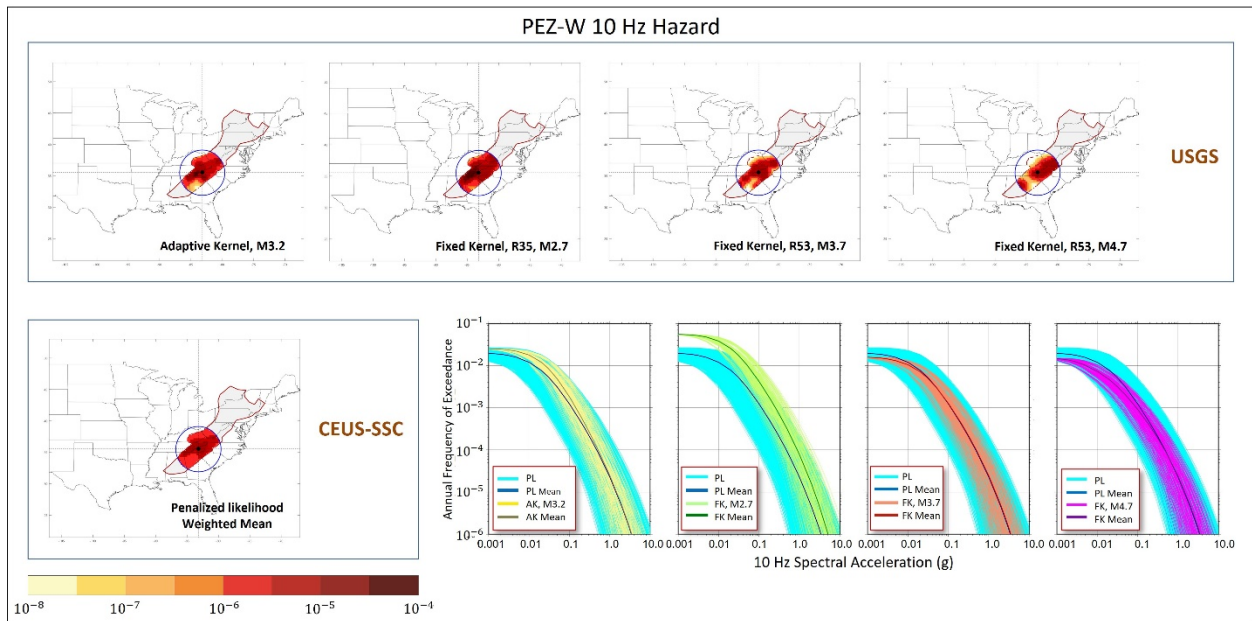


Figure 4-14: Hazard curves computed at 10 Hz spectral acceleration at a select site in the Paleozoic Extended Crust-Wide (PEZ-W) seismotectonic zone (see Table 4-2 for coordinates) using Adaptive Kernel (AK), Fixed Kernel (FK), and Penalized-Likelihood (PL) methods. Maps show results computed with different approaches.

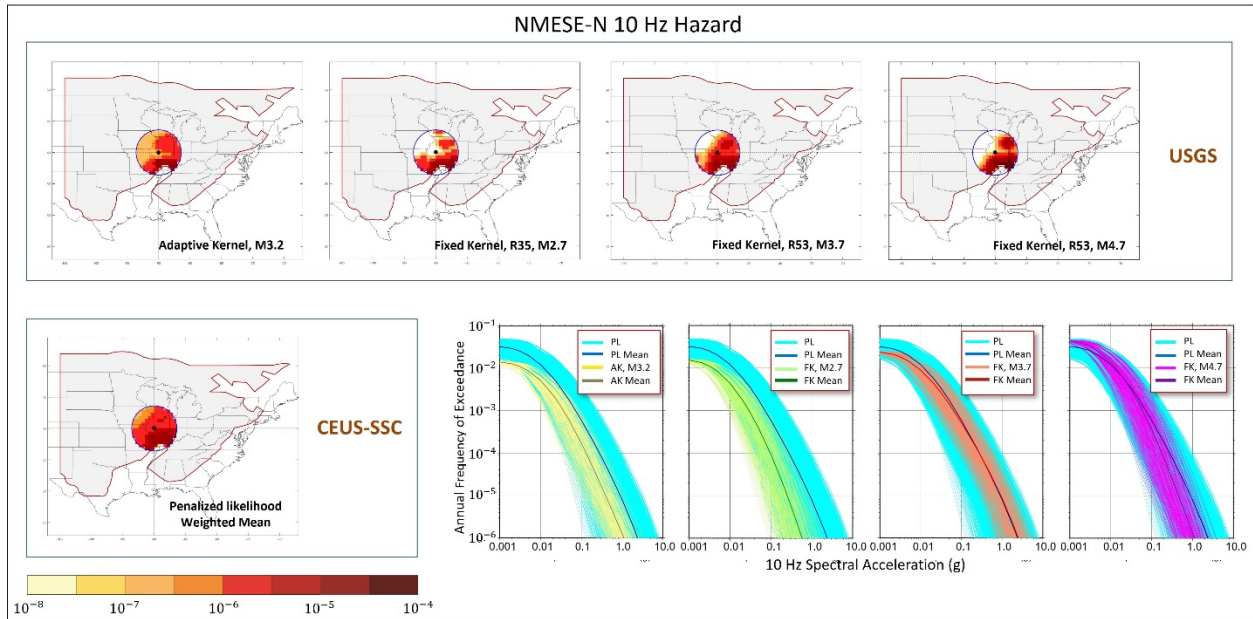


Figure 4-15: Hazard curves computed at 10 Hz spectral acceleration at a select site in Non-Mesozoic-and-Younger Extension-Narrow (NMESE-N) seismotectonic zone (see Table 4-2 for coordinates) using Adaptive Kernel (AK), Fixed Kernel (FK), and Penalized-Likelihood (PL) methods. Maps show results computed with different approaches.

Figure 4-16 compares different combinations of the mean and weighted mean hazard curves at select sites in NAP, PEZ-W, and NMESE-N seismotectonic zones at 10 Hz. Focusing on the mean Penalized-Likelihood (PL) and the weighted kernel (adaptive and fixed) hazard curves (frames on the right), they are very similar for NAP, specially at spectral accelerations greater than $0.05 g$. For PEZ-W, the PL mean hazard curve is lower than the weighted kernel curve at all spectral accelerations, whereas the opposite is true for the NMESE-N zone. This indicates that the distributed seismicity rates estimated by different smoothing methods are affected by the initial inputs (earthquake catalog). This is evidenced by the total rates for $M > 5.0$ calculated for each smoothing method (charts in Figure 4-16). For example, fixed kernel (FK) total rate estimates with catalog completeness above M2.7 (FK_M2.7) in PEZ-W and FK_M4.7 in NMESE-N zones are much larger than those calculated from other methods in the same zones.

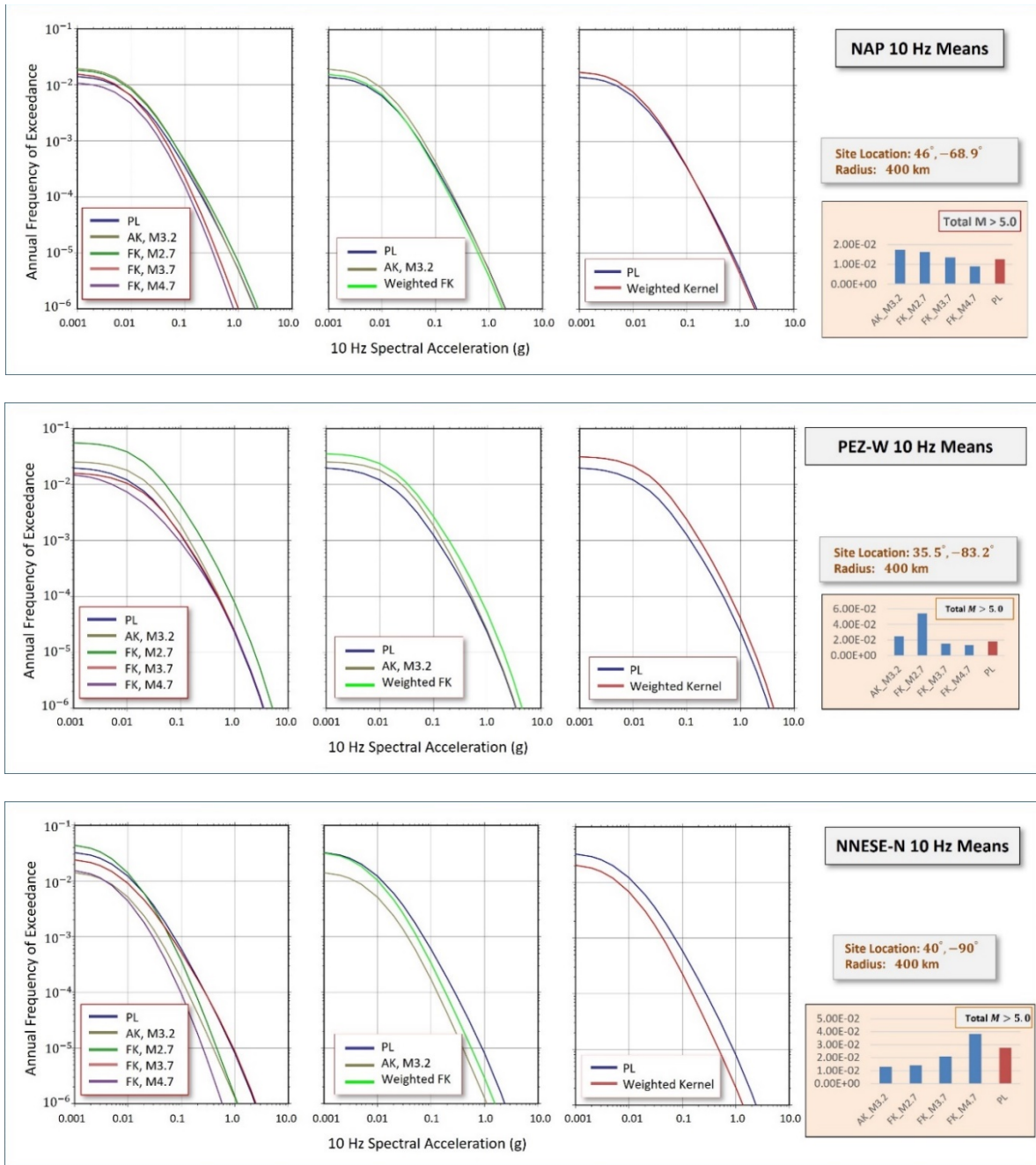


Figure 4-16: Various combinations of the mean and weighted mean hazard curves computed at 10 Hz spectral acceleration at select sites in North Appalachian (NAP), Paleozoic Extended Crust-Wide (PEZ-W), and Non-Mesozoic-and-Younger Extension-Narrow (NNESE-N) seismotectonic zones are plotted here. The frames on the left show the mean hazard curves for Penalized-Likelihood (PL), Adaptive Kernel (AK), and three Fixed Kernel (FK) methods. The middle frames show PL mean (average of 24 rate realizations), AK mean, and the weighted mean of FK (average of three different FK approaches). The frames on the right show the PL mean and the kernel weighted mean ($0.4 \times AK + 0.6 \times FK$) hazards.

5 EVALUATIONS OF EPISTEMIC UNCERTAINTY

Evaluating epistemic uncertainty is an important component of a seismic hazard analysis. One of the key concepts described in the Senior Seismic Hazard Analysis Committee (SSHAC) process developed in 1997 (NRC, 1997) is establishing the center, body, and range of technically defensible interpretations of data and models that affect the seismic hazard analysis. This process was used in the development of the Central and Eastern United States Seismic Source Characterization (CEUS-SSC) study (NRC, 2012). The center, body, and range of background seismicity rates in the Central and Eastern United States (CEUS) is thought to be captured by using the penalized-likelihood approach described in Section 3.2, having three cases, Cases A, B, and E, which have differing minimum magnitudes and different weighting of magnitude bins. In addition, eight random gridded realizations of seismicity rates are developed for each case. One of the objectives of this study on background seismicity rates is to independently evaluate how well the penalized-likelihood approach captures the center, body, and range of background seismicity rates in the CEUS.

The adequacy of the penalized-likelihood approach as implemented in the CEUS-SSC for capturing the center, body, and range of background seismicity rates was evaluated by comparing rates obtained from the penalized-likelihood approach with background seismicity rates obtained using fixed and adaptive kernel approaches that are implemented by the USGS for developing the National Seismic Hazard Model (NSHM) and by comparing seismic hazard at three sites using the various background seismicity rate approaches. Background seismicity and hazard curve comparisons are presented in Section 4. Some figures are repeated here. Figure 4-1 shows moderate differences in the distribution of background seismicity between fixed kernel and other smoothing approaches. However, the total rates for the seismotectonic zones are generally similar, with the exception of the fixed kernel rates when using a minimum magnitude of 2.7 (Figure 5-1). In Paleozoic Extended Crust-Wide (PEZ-W), the fixed kernel with a minimum magnitude of 2.7 produces a total seismicity rate approximately two times greater than all other kernel and penalized-likelihood rates. The total seismicity rate for the Non-Mesozoic-and-Younger Extension-Narrow (NMESE-N) M_{max} zone is also moderately higher for the fixed kernel with a minimum magnitude of 2.7, and this is likely the result of increased rates in the PEZ-W seismotectonic zone, which is encapsulated within the NMESE-N.

When comparing the hazard curves, one observes in Figure 5-2 through Figure 5-4 that the hazard curves associated with the penalized-likelihood rates generally encapsulate the hazard curves obtained using either fixed or adaptive kernel smoothed rates for exceedance frequencies of interest in developing a ground motion response spectrum (GMRS) for large light water reactors (annual exceedance frequencies of 10^{-4} to 10^{-5}). Overall, the hazard curve comparison provides confidence that the penalized-likelihood approach as implemented in the CEUS-SSC reasonably captures the center, body, and range of technically defensible interpretations of background seismicity rates within the CEUS for large light water reactor applications.

Mean hazard curves are compared in Figure 5-5. When weights used by the USGS for combining mean fixed kernel results are applied, the mean hazard for the weighted fixed kernel is within the range of hazard obtained using the adaptive kernel and penalized-likelihood rates for NAP and NMESE-N and higher for PEZ-W. A question raised by differences observed between the weighted kernel and penalized-likelihood mean hazard is whether kernel approaches should be used in addition to the penalized-likelihood approach to evaluate the mean hazard in the CEUS.

At higher annual frequencies of exceedance (4×10^{-3} to 4×10^{-4}), the hazard obtained using the penalized-likelihood rates may not capture the range of hazard obtained using the kernel smoothing approaches. This is most pronounced for the fixed kernel method with a minimum magnitude of 2.7. This could have implications for advanced or micro-reactors that use the forthcoming 10 CFR Part 53 for licensing where the design basis ground motion or safe shutdown earthquake may have a higher annual frequency of exceedance than is considered in the design of large light water reactors. Additional work may be needed to assess how defensible the fixed kernel results are in the PEZ-W seismotectonic zone when using a minimum magnitude of 2.7.

The comparisons of this study, between and among the USGS and CEUS-SSC rate forecasts, suggest potential paths to increasing and improving rate forecasts in the CEUS. The smoothing approaches presented in this RIL for seismicity-based earthquake rate forecasts couple the total seismicity rate within a region (e.g., $M_W \geq 5$) with the spatial distribution of these rates. These forecast parameters have been decoupled in the USGS forecasts for California (i.e., Uniform California Earthquake Rupture Forecast, Version 3 [UCERF3]; Field et al., 2014). We find significant differences in total regional seismicity rates (Figure 4-1, Figure 4-2, and Figure 4-3) that suggest potential for improved characterization of epistemic uncertainty in the use of smoothed seismicity models. Estimating spatial rate, however, inherently relies on moment magnitude of small earthquakes which may include artifacts from conversion relationships that propagate into seismic hazard calculations. This effect is likely amplified near the boundaries where regional seismic networks produce authoritative source estimates. Encouraging regional networks to harmonize magnitude estimates may reduce some of these effects.

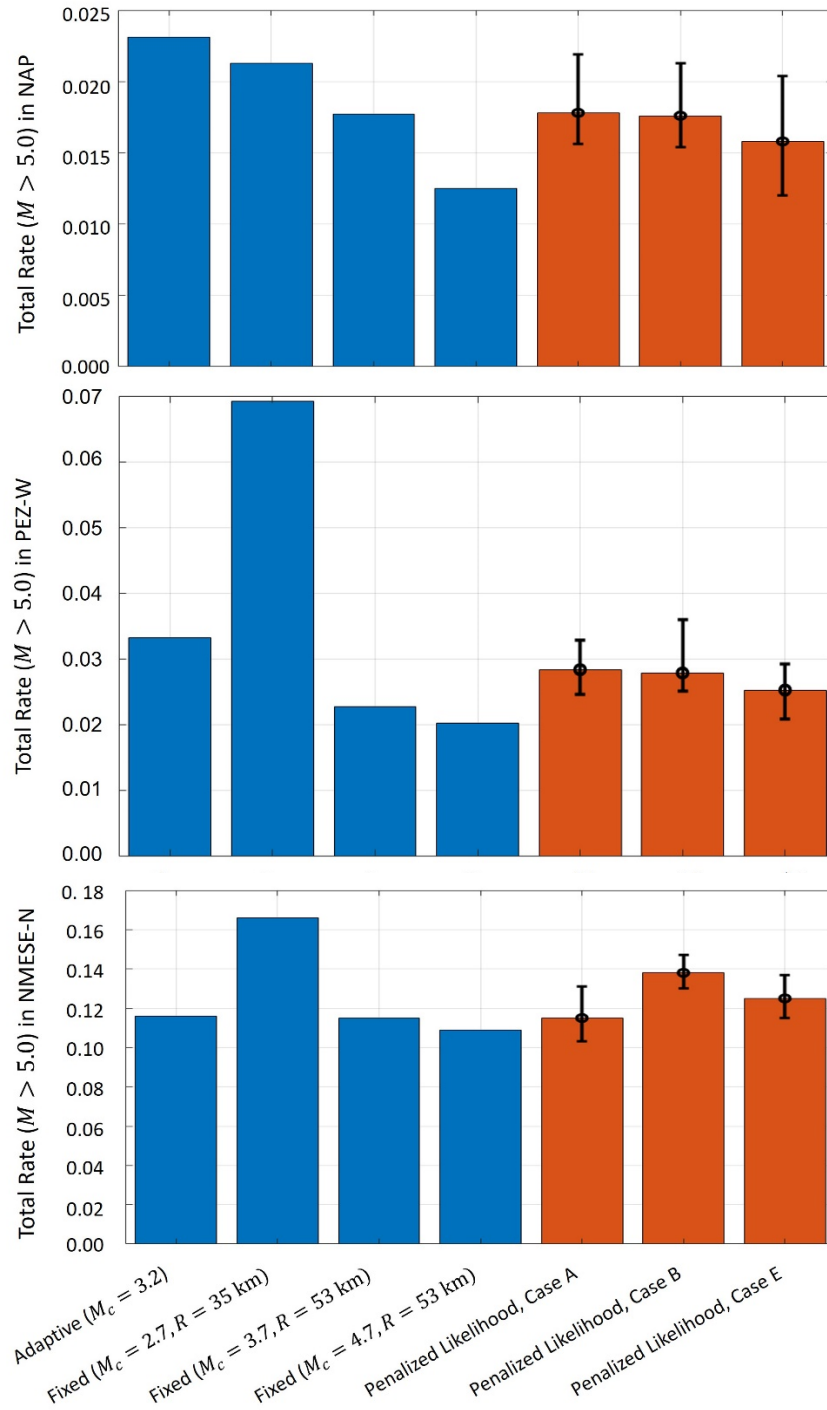


Figure 5-1: Total seismicity rates for North Appalachian (NAP), Paleozoic Extended Crust-Wide (PEZ-W), and Non-Mesozoic-and-Younger Extension-Narrow (NMESE-N) seismotectonic zones. The error bars indicate the range of total seismicity rates for eight realizations in each of the three cases (A, B, and E).

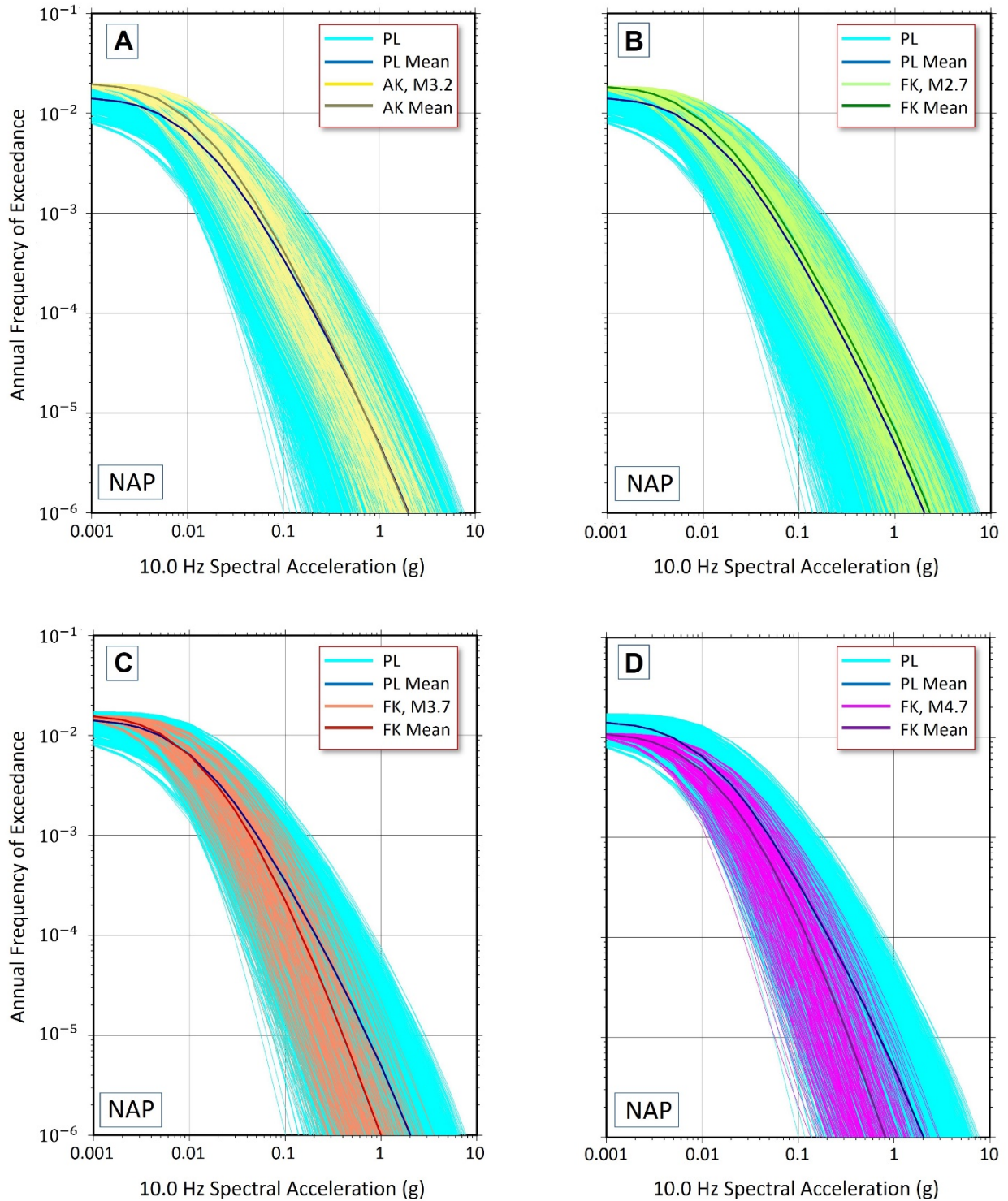


Figure 5-2: North Appalachian (NAP) 10 Hz hazard curves comparing penalized-likelihood (PL) results with results using (A) adaptive kernel (AK), (B) fixed kernel (FK) with a minimum magnitude of 2.7, (C) fixed kernel with a minimum magnitude of 3.7, and (D) fixed kernel with a minimum magnitude of 4.7.

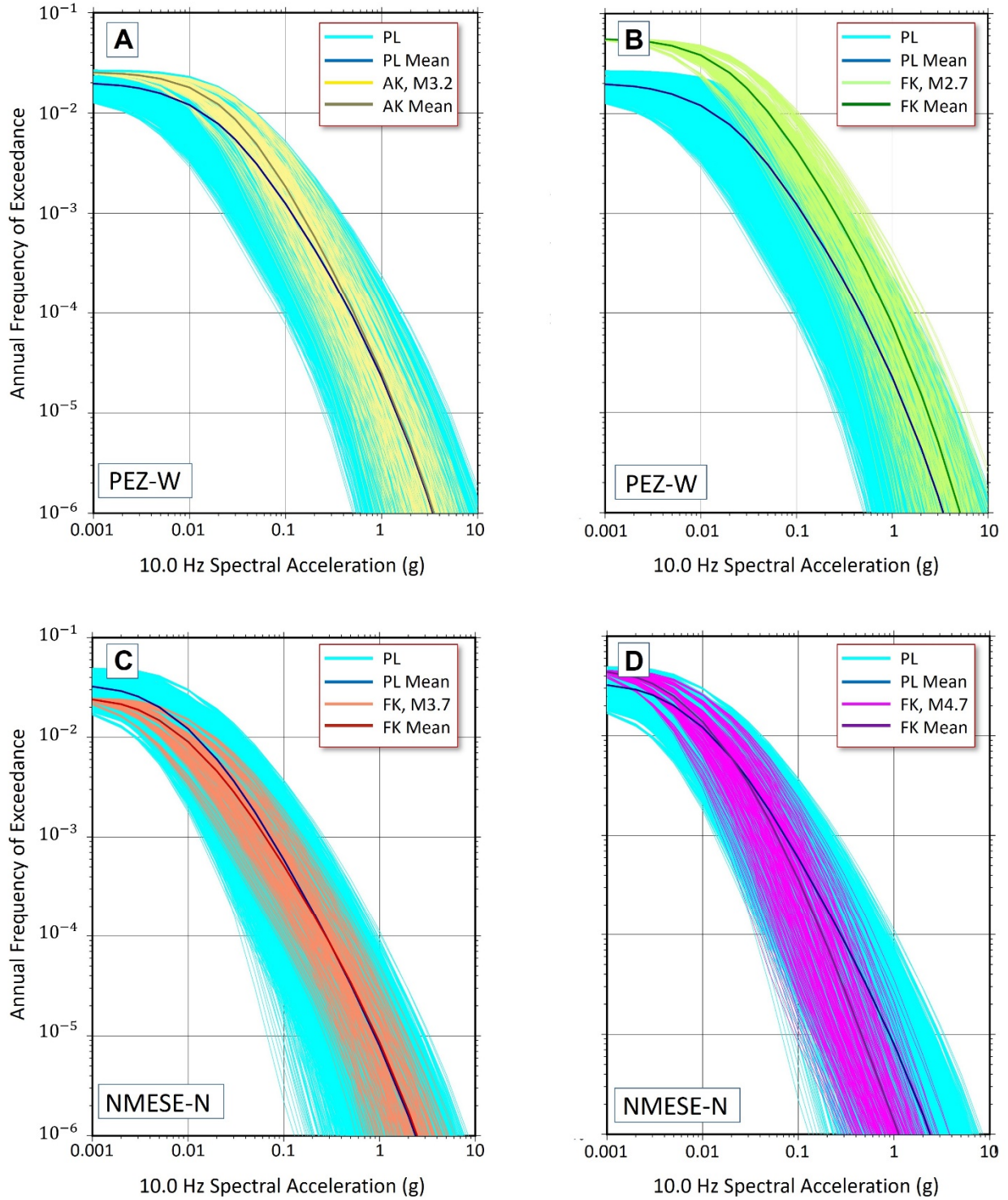


Figure 5-3: Paleozoic Extended Crust-Wide (PEZ-W) 10 Hz hazard curves comparing penalized-likelihood (PL) results with results using (A) adaptive kernel (AK), (B) fixed kernel (FK) with a minimum magnitude of 2.7, (C) fixed kernel with a minimum magnitude of 3.7, and (D) fixed kernel with a minimum magnitude of 4.7.

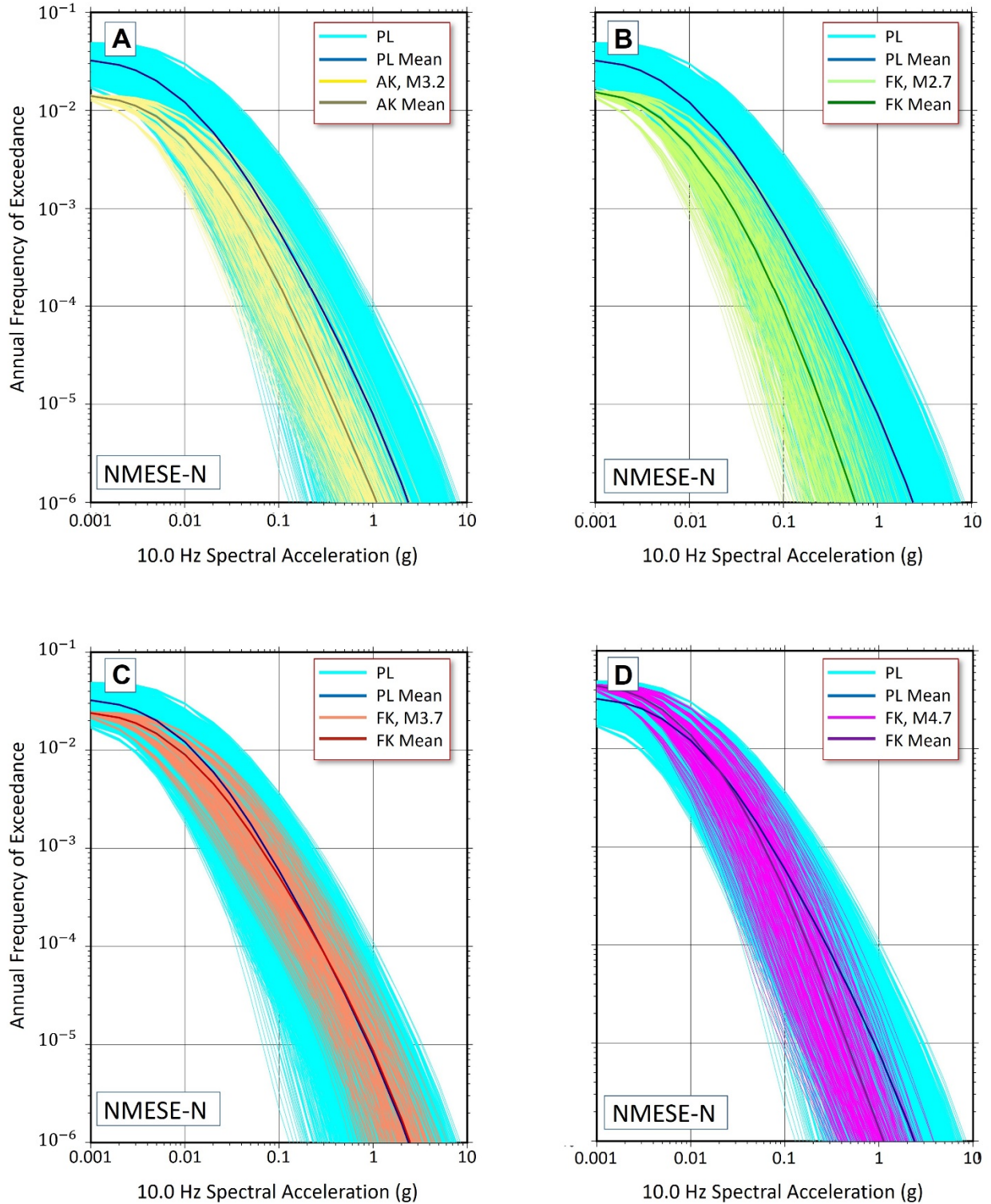


Figure 5-4: Non-Mesozoic-and-Younger Extension-Narrow (NMESE-N) 10 Hz hazard curves comparing penalized-likelihood (PL) results with results using (A) adaptive kernel (AK), (B) fixed kernel (FK) with a minimum magnitude of 2.7, (C) fixed kernel with a minimum magnitude of 3.7, and (D) fixed kernel with a minimum magnitude of 4.7.

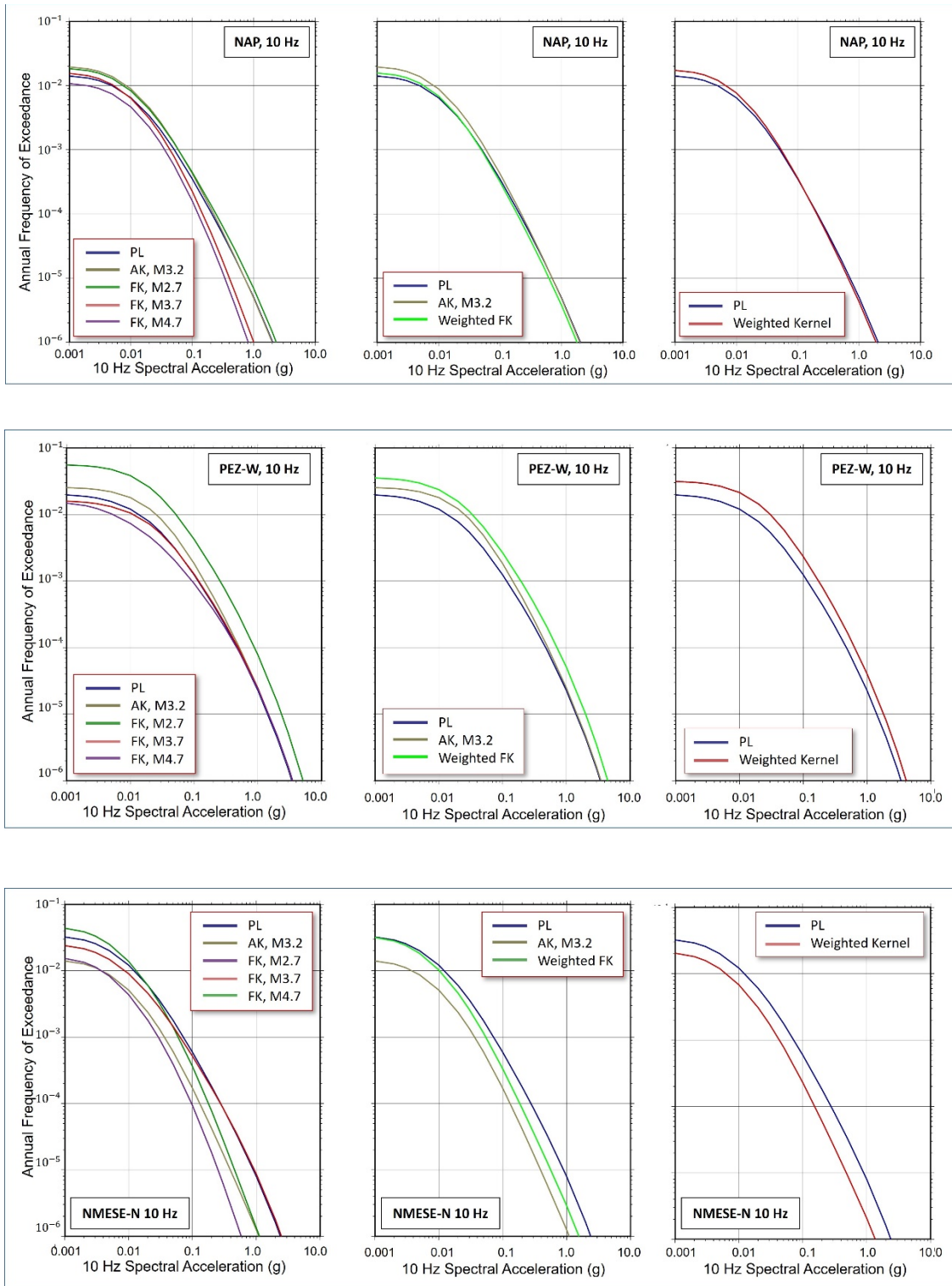


Figure 5-5: The mean and weighted mean hazard curves at select sites in North Appalachian (NAP), Paleozoic Extended Crust-Wide (PEZ-W), and Non-Mesozoic-and-Younger Extension-Narrow (NMESE-N) seismotectonic zones are plotted here for Penalized-Likelihood (PL), Adaptive Kernel (AK), and Fixed Kernel (FK) methods.

6 CONCLUSIONS

This study was performed in part because previous studies indicated that background seismicity may be the cause of differences between USGS and CEUS-SSC mean hazard (Petersen et al., 2018). Background seismicity rates are affected by models used to convert small magnitude earthquakes to moment magnitude. Seismicity rates and the distribution of seismicity rates are also affected by the models used to smooth background seismicity. This study demonstrated results from coda envelop calibration for estimating the moment magnitude of small earthquakes, provided an in-depth review of common approaches used in practice to smooth background seismicity rates, and compared seismicity rates and hazard using those approaches. Conclusions and recommendations based on this work, and recommendations for future work on magnitude conversion and background seismicity smoothing, are provided below.

The inconsistencies and bias in magnitude conversion for small magnitude events has broad implications for inferred seismicity rates (a -values) and b -values. Magnitudes estimated for a given earthquake may vary substantially depending on magnitude type, details of calculation procedure, and geologic structure. Inconsistencies in magnitude conversion procedures were illustrated by large differences between CEUS-SSC converted and coda envelope-based magnitudes in southern Kansas and northern Oklahoma.

Magnitude conversion relations in current practice perform increasingly poorly at smaller magnitudes in the regions of Oklahoma and Kansas where data were available to apply the coda envelop-based conversion approach. This is likely because the conversion relations were developed with few (or no) reliable events at small magnitudes. Likewise, b -values are often determined primarily by earthquakes within a size range encompassing transitions among magnitude types, and associated biases cause potential for misinterpretation or large errors in projected event rates at higher magnitudes.

Avenues to mitigate the current effects of magnitude inconsistencies include the following:

1. Acknowledgement that systematic magnitude biases (not just Gaussian uncertainty) can be large and accounting for these potential biases in related interpretations. Discounting these biases can lead to unrealistic conclusions.
2. Targeted studies to develop magnitude conversion relationships with lower bias, particularly focused on small magnitudes. Conversion relationships can only be expected to be valid if they are derived using a dataset from the same geographic area, magnitude calculation procedure (which depends both on the monitoring agency and time period), and magnitude range to which they are applied.
3. Exploration of operational estimation for events currently lacking routine magnitudes. Although challenging, this represents a worthy future goal for seismic monitoring agencies. Coda envelope analysis similar to that applied here may provide a path toward this goal. Future software developments to improve automatic windowing for coda measurements could reduce the need for manual intervention and facilitate large-scale processing. Magnitude uncertainty quantification is another area for future development – robust uncertainties could further improve the ability to use these magnitudes in rigorous analyses.

In the meantime, users should approach magnitudes of small earthquakes and associated conversion relationships with appropriate caution.

Penalized-likelihood and kernel smoothing approaches were compared within three different zones in the CEUS, and hazard from background seismicity was computed at a site within each zone using background seismicity rates produced by these approaches. The comparison of smoothed seismicity rates showed that the penalized-likelihood and adaptive kernel smoothing approaches are similar in both the distribution of seismicity and the total seismicity rate for a seismotectonic or M_{max} zone. Although the distribution of seismicity rates is generally similar, differences in the distribution may result in differences in mean hazard when using these two approaches. The distribution of seismicity rates for the fixed kernel smoothing approach is substantially different, especially at distances farther from locations of recorded seismic events due to the rapid decay of seismic rates at distances greater than twice the kernel width. The fixed kernel smoothing results presented in this report did not include a minimum or floor rate, which can contribute to some of the differences at low fixed kernel seismicity rates.

The comparison of hazard curves in the 10^{-4} to 10^{-5} annual frequency of exceedance range indicates that the penalized-likelihood approach as implemented in the CEUS-SSC SSHAC study does capture the center, body, and range of seismicity rate smoothing models. Differences in the penalized-likelihood and weighted kernel mean hazard were observed. Future updates to the CEUS-SSC may consider whether adaptive and fixed kernel smoothing approaches should be explicitly incorporated into the SSC logic tree to modify the mean hazard and, if incorporated, how much weight should be assigned to the kernel models. Explicitly incorporating the kernel-based background seismicity rates would reduce differences between USGS and CEUS-SSC derived hazard. The drawback of adding the kernel approaches is increased logic tree complexity and computational effort for developing hazard curves. With adoption the NGA-East GMM for assessing hazard in the CEUS, additional computational effort is not trivial.

7 REFERENCES

- Al-Ismaïl, F., Ellsworth, W. L., and Beroza, G. C. (2020). Empirical and synthetic approaches to the calibration of the local magnitude scale, M_L , in southern Kansas. *Bulletin of the Seismological Society of America*, 110(2), 689-697, doi: 10.1785/0120190189.
- Barno, J. (2017). LLNL/coda-calibration-tool. *U.S. Department of Energy, Lawrence Livermore National Laboratory*, doi: 10.11578/dc.20180306.1.
- Darold, A., Holland, A. A., Chen, C., and Youngblood, A. (2014). Preliminary analysis of seismicity near Eagleton 1–29, Carter County, July 2014. *Oklahoma Geol. Survey Open-File Rept. OF2-2014*, 16 pp., available at <http://ogs.ou.edu/docs/openfile/OF2-2014.pdf> (last accessed November 2020).
- Electric Power Research Institute-Seismicity Owners Group (EPRI-SOG) (1986). Seismic hazard methodology for the central and eastern United States: Volumes 5-10, *EPRI Publication NP-44726-V5*, 412 pp.
- Electric Power Research Institute (EPRI) (1988). Seismic hazard methodology for the central and eastern United States: 10 volumes, *EPRI Publication NP-4726*.
- Electric Power Research Institute/Department of Energy/Nuclear Regulatory Commission (EPRI/DOE/NRC) (2012). Central and eastern United States seismic source characterization for nuclear facilities, Final Report. *EPRI, USDOE, and USNRC*, Palo Alto, California, available at www.ceus-ssc.com (last accessed November 2020)
- Frankel, A. D., Mueller, C. S., Barnhard, T. P., Perkins, D. M., Leyendecker, E. V., Dickman, N., ... and Hopper, M. G. (1996). National Seismic-Hazard Maps: Documentation June 1996. *U.S. Geological Survey Open-File Report 96-532*, 110 pp., doi: 10.3133/ofr96532.
- Gardner, J. K., and Knopoff, L. (1974). Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian? *Bulletin of the Seismological Society of America*, 64(5), 1363-1367, doi: 10.1785/BSSA0640051363.
- Goulet, C. A., Bozorgnia, Y., Abrahamson, N. A., Kuehn, N., Al Atik, L., Youngs, R., and Graves, R. (2018). Central and eastern North America ground-motion characterization—NGA-East final report. *Pacific Earthquake Engineering Research Report No. 2018/08*, 817 pp.
- Greig, D., Yenier, E., Baturan, D., and Karimi, S. (2018). Determination of local magnitude distance corrections for northern Oklahoma. *Seismological Research Letters*, 89(5), 1786-1795, doi: 10.1785/0220170225.
- Helmstetter, A., Kagan, Y. Y., and Jackson, D. D. (2007). High-resolution time-independent grid-based forecast for $M \geq 5$ earthquakes in California. *Seismological Research Letters*, 78(1), 78-86, doi: 10.1785/gssrl.78.1.78.
- Herrmann, R. B., Benz, H., and Ammon, C. J. (2011). Monitoring the earthquake source process in North America. *Bulletin of the Seismological Society of America*, 101(6), 2609-2625, doi: 10.1785/0120110095.
- Kafka, A. L. (2007). Does seismicity delineate zones where future large earthquakes are likely to occur in intraplate environments? in Stein, S., and Mazzotti, S. (editors), *Continental Intraplate Earthquakes: Science, Hazard, and Policy Issues*, *Geological Society of America Special Paper 425*, 35-48, doi: 10.1130/2007.2425(03).
- Kafka, A. L. (2009). Use of seismicity to define seismic sources: Application to eastern North America. *Presentation given at CEUS SSC Project Workshop #2*, February 18-20, Palo Alto, Calif.
- Mayeda, K., Hofstetter, A., O'Boyle, J. L., and Walter, W. R. (2003). Stable and transportable regional magnitudes based on coda-derived moment-rate spectra. *Bulletin of the Seismological Society of America*, 93(1), 224-239, doi: 10.1785/0120020020.
- Moschetti, M. P., Powers, P. M., Petersen, M. D., Boyd, O. S., Chen, R., Field, E. H., ... and Zeng, Y. (2015). Seismic source characterization for the 2014 update of the U.S. National

- Seismic Hazard Model. *Earthquake Spectra*, 31(1_suppl), S31-S57, doi: 10.1193/110514EQS183M.
- Mueller, C. S. (2019). Earthquake catalogs for the USGS national seismic hazard maps. *Seismological Research Letters*, 90(1), 251-261, doi: 10.1785/0220170108.
- Mueller, C.S. (2023). Earthquake catalog (1568 to 2018) for the USGS National Seismic Hazard Model and Nuclear Regulatory Commission: U.S. Geological Survey data release, <https://doi.org/10.5066/P95SNP2J>.
- NRC (1997). Recommendations for probabilistic seismic hazard analysis: Guidance on uncertainty and use of experts. *U.S. Nuclear Regulatory Commission NUREG/CR 6372*, Washington, D.C. (<https://www.nrc.gov/docs/ML0800/ML080090003.pdf>)
- NRC (2012). Central and eastern United States seismic source characterization for nuclear facilities. *NUREG-2115*, vol. 2. (<http://www.ceus-ssc.com>)
- Nuttli, O. W. (1973). Seismic wave attenuation and magnitude relations for eastern North America. *Journal of Geophysical Research*, 78(5), 876-885, doi: 10.1029/JB078i005p00876.
- Nuttli, O. W. (1986). Yield estimates of Nevada Test Site explosions obtained from seismic Lg waves. *Journal of Geophysical Research: Solid Earth*, 91(B2), 2137-2151, doi: 10.1029/JB091iB02p02137.
- Petersen, M. D., Moschetti, M. P., Powers, P. M., Mueller, C. S., Haller, K. M., Frankel, A. D. ... and Olsen, A. H. (2014). Documentation for the 2014 update of the United States National Seismic Hazard Maps, *U.S. Geological Survey Open-File Report 2014-1091*, 243 pp., doi: 10.3133/ofr20141091.
- Petersen, M. D., Boyd, O. S., Hartzel, S., McNamara, D., Moschetti, M. P., Mueller, C. S., ... and Shumway, A. (2018). Uncertainties in seismic hazard assessment, *U.S. NRC ADAMS ML18240A369*. (<https://adamswebsearch2.nrc.gov/webSearch2/main.jsp?AccessionNumber=ML18240A369>)
- Petersen, M. D., Shumway, A. M., Powers, P. M., Moschetti, M. P., Llenos, A. L., Michael, A. J., ... and Shiro, B. R. (2022). 2021 US National Seismic Hazard Model for the State of Hawaii. *Earthquake Spectra*, 38(2), 865-916, doi: 10.1177/875529302110520.
- Petersen, M. D., Mueller, C. S., Moschetti, M. P., Hoover, S. M., Llenos, A. L., Ellsworth, W. L., ... and Rukstales, K. S. (2016). Seismic-hazard forecast for 2016 including induced and natural earthquakes in the central and eastern United States. *Seismological Research Letters*, 87(6), 1327-1341, doi: 10.1785/0220160072.
- Petersen, M.D., Mueller, C.S., Moschetti, M.P., Hoover, S.M., Rubinstein, J.L., Llenos, A.L., ... and Anderson, J.G. (2015). Incorporating induced seismicity in the 2014 United States National Seismic Hazard Model—Results of 2014 workshop and sensitivity studies. *U.S. Geological Survey Open-File Report 2015-1070*, 69 pp., doi: 10.3133/ofr20151070.
- Petersen, M. D., Shumway, A. M., Powers, P. M., Mueller, C. S., Moschetti, M. P., Frankel, A. D., ... and Zeng, Y. (2020). The 2018 update of the US National Seismic Hazard Model: Overview of model and implications. *Earthquake Spectra*, 36(1), 5-41, doi: 10.1177/8755293019878199.
- Rigsby, C., Herrmann, R. B., and Benz, H. (2014). An investigation of MbLg versus Mw for eastern North America. *Seismological Research Letters*, 85(3), 625-630, doi: 10.1785/0220130138.
- Rubinstein, J. L., Ellsworth, W. L., and 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.
- Shelly, D. R., Mayeda, K., Barno, J., Whidden, K. M., Moschetti, M. P., Llenos, A. L., ... and Walter, W. R. (2021). A big problem for small earthquakes: Benchmarking routine magnitudes and conversion relationships with coda envelope-derived Mw in southern

- Kansas and northern Oklahoma. *Bulletin of the Seismological Society of America*, 112(1), 210-225, doi: 10.1785/0120210115.
- Shi, Y., and Bolt, B. A. (1982). The standard error of the magnitude-frequency b value. *Bulletin of the Seismological Society of America*, 72(5), 1677-1687, doi: 10.1785/BSSA0720051677.
- Stepp, J.C. (1972). Analysis of completeness of the earthquake sample in the Puget Sound area and its effect on statistical estimates of earthquake hazard. *Proceedings of the International Conference on Microzonation*, 2, 897-910.
- Utsu, T. (1966). A statistical significance test of the difference in b-value between two earthquake groups. *Journal of Physics of the Earth*, 14(2), 37-40, doi: 10.4294/jpe1952.14.37.
- Walter, J. I., Ogwari, P., Thiel, A., Ferrer, F., Woelfel, I., Chang, J. C., ... and Holland, A. A. (2020). The Oklahoma Geological Survey Statewide Seismic Network. *Seismological Research Letters*, 91(2A), 611-621, doi: 10.1785/0220190211.
- Weichert, D. H (1980). Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes. *Bulletin of the Seismological Society of America*, 70(4), 1337-1346, doi: 10.1785/BSSA0700041337.