

Cal Poly

Caltech



UC Irvine

UCLA

**UC Santa
Barbara**

USC

Surface Fault Displacement Models for Strike-Slip Faults

Brian S.J. Chiou, Ph.D.

California Department of Transportation
Sacramento, California

Rui Chen, Ph.D., P.G.

Kate Thomas

California Geological Survey
Sacramento, California

Christopher W.D. Milliner, Ph.D.

California Institute of Technology
Pasadena, California

Timothy E. Dawson, P.G., C.E.G.

California Geological Survey
Sacramento, California

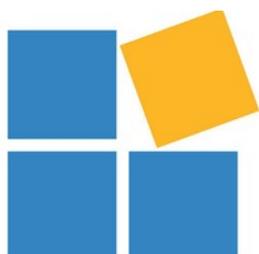
Mark D. Petersen, Ph.D., P.G., C.E.G.

United States Geological Survey
Golden, Colorado

A report on research conducted with support from California
Department of Transportation, California Geological Survey,
and California Energy Commission.

Report GIRS-2022-07
DOI: 10.34948/N3RG6X

University of California, Los Angeles (headquarters)



Natural Hazards Risk & Resiliency Research Center

B. John Garrick Institute for the Risk Sciences

Surface Fault Displacement Models for Strike-Slip Faults

Brian S.J. Chiou, Ph.D.

California Department of Transportation, Sacramento, CA

Rui Chen, Ph.D., P.G.

Kate Thomas

California Geological Survey, Sacramento, CA

Christopher W.D. Milliner, Ph.D.

California Institute of Technology, Pasadena, CA

Timothy E. Dawson, P.G.

California Geological Survey, Sacramento, CA

Mark D. Petersen, Ph.D.

U.S. Geological Survey, Golden, CO

A report on research conducted with support from California Department of Transportation, California Geological Survey, and California Energy Commission.

Report GIRS-2022-07

Natural Hazards Risk and Resiliency Research Center
B. John Garrick Institute for the Risk Sciences
University of California, Los Angeles (Headquarters)

April 2023

DISCLAIMER:

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government or California State Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Chiou, B.S.J., Chen, R., Thomas, K., Milliner, C.W.D., Dawson, T.E., and Petersen, M.D., 2023, Surface Fault Displacement Models for Strike-Slip Faults, Report GIRS-2022-07, University of California, Los Angeles, 186 p., <https://doi.org/10.34948/N3RG6X>.

ABSTRACT

Fault displacement models (FDMs) are an essential component of the probabilistic fault displacement hazard analyses (PFDHA), much like ground motion models in the probabilistic seismic hazard analyses for ground motion hazards. In this study, we develop several principal surface FDMs for strike-slip earthquakes. The model development is based on analyses of the new and comprehensive fault displacement database developed as part of the Fault Displacement Hazard Initiative project led by the University of California, Los Angeles. The main objective of our study is to update the FDMs that were developed over a decade ago by the U.S. Geological Survey and California Geological Survey, in which a reference trace was drawn manually, FDMs are fixed-effect models for lateral displacement, displacement on multiple subparallel ruptures is not aggregated, magnitude (M) scaling is linear, and natural logarithm of displacement is assumed to be normally distributed. In the current study, the net displacement data for each selected earthquake are analyzed in a local coordinate system that tracks a main rupture trace developed using a semi-automatic approach based on the Least-Cost Pass analysis. Displacements across multiple subparallel principal rupture traces are aggregated at each along-strike measurement location. In addition to updating the displacement data, we have included model formulation updates such as a bilinear M scaling, random-intercept mixed-effect modeling and M -dependent variance of the random intercept, and non-normal probability distribution of the natural logarithm of aggregated principal net displacement. Our preferred model assumes a negative exponentially modified Gaussian (nEMG) distribution, and it performs well at representing variability in fault displacement data. These updates substantially improve data fit and provide reasonable quantile predictions, particularly for the upper quantiles of large magnitude ruptures. The nEMG distribution affords analytic expressions of probability density function and cumulative distribution function, making its implementation in PFDHA straightforward, without resorting to numerical integration or stochastic simulation. Furthermore, the stochastic representation of the nEMG random variate is interpreted as due to the occurrence of interior tapering in surface slip, making it a fitting model statistically as well as physically. Several interim updates assuming different probability distributions lead progressively to the final and preferred model. Each of these interim models can potentially serve as an alternative FDM in PFDHA to model the epistemic uncertainty in the choice of probability distribution. Applications of the updated models are demonstrated using PFDHA examples. The steeper slope of the resulting hazard curve from the non-normal distribution models, particularly the nEMG model, yield displacement at a low exceedance rate that is much smaller than traditional normal-distribution-based models.

ACKNOWLEDGMENTS

Support for this project was provided by California Department of Transportation and California Geological Survey (CGS) for staff time and partially by funding awarded to CGS by University of California, Los Angeles (UCLA) as part of a larger project funded by California Energy Commission.

The support of these organizations is gratefully appreciated. The opinions, findings, conclusions, and recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the study sponsors, the B. John Garrick Risk Institute, or the Regents of the University of California.

The authors benefited from frequent interactions and constructive discussions with other researchers, hazard analysts, and practitioners participating in the Fault Displacement Hazard Initiative (FDHI) project led by UCLA. We thank Professor Yousef Bozorgnia for initiating and leading the FDHI project and for review of this report, Steve Thompson for offering inspiring comments and suggestions, Alexandra Sarmiento for help with the FDHI database and review of this report, and Yongfei Wong for sharing his simulation results that provided valuable comparison and guidance in building magnitude-scaling relation in our models. We are grateful to Donald Wells for generously providing his data to help us evaluate the applicable magnitude range of our models. We are indebted to Neal (Simon) S. Kwong, Alexandra E. Hatem, Brian R. Shiro, and Janet M. Carter of U.S. Geological Survey. Their thorough reviews, constructive comments, and effective editing helped to improve the clarity and quality of this report.

CONTENTS

ABSTRACT	iii
ACKNOWLEDGMENTS	iv
CONTENTS	v
LIST OF FIGURES	viii
LIST OF TABLES	xviii
1 Introduction	1
2 Rupture and Displacement Data	4
2.1 FDHI database.....	4
2.2 Data selection	6
2.2.1 Excluded Events and Rationale for Exclusion.....	7
2.2.2 Dataset Selection.....	8
2.2.3 Excluded Ruptures and Measurements in Selected Datasets.....	9
2.2.4 Modification to FDHI Principal and Distributed Classification.....	10
2.3 Main rupture trace and coordinate system.....	11
2.3.1 Least-Cost Path Analysis	13
2.3.2 LCP as Main Rupture Trace and GC2 as Coordinate System	13
2.3.3 Comparison of LCP with ECS.....	14
2.3.4 Significance of LCP as Main Rupture Trace	15
2.4 Aggregated net principal displacement.....	16
2.4.1 Simple Segmentation Models	17
2.4.2 Computation of Aggregated Principal Displacement	17
2.4.3 Comparison of Aggregated and Measured Displacement Profiles.....	18
2.5 Data distribution	18
3 Development of Probabilistic Distribution Models for Principal Displacement.....	31
3.1 Response variable and its probability distribution.....	32
3.2 Interim updates.....	33
3.2.1 <i>Simple Update</i> : Repeating P11 Regression Analysis on Selected FDHI Data	34
3.2.2 <i>Model1.NO</i> : Mixed-Effect Regression of Tier-1 Data	37
3.2.3 <i>Model2.SN</i> : Distributional Regression Using Skew-Normal Distribution	42
3.2.4 <i>Model3.ST</i> : Distributional Regression Using Skew- <i>t</i> Distribution.....	47

3.3	<i>Model4.nEMG</i> : distributional regression using Negative Exponentially Modified Gaussian distribution	49
3.3.1	Stochastic Representation of $\ln(D)$	49
3.3.2	<i>nEMG Distribution</i>	50
3.3.3	Models for μ, σ , and v	51
3.3.4	Variance Model of the Random Intercept of μ	52
3.3.5	Diagnostics of Model4.nEMG, Normalized Quantile Residual	52
3.3.6	Predictive Distribution	52
3.4	Comparison of interim and preferred models	53
3.5	Assessment of estimation uncertainty	54
3.5.1	Uncertainty in M-Scaling.....	54
3.5.2	Uncertainty of Other Coefficients.....	57
3.6	Discussions.....	57
3.6.1	Predicted Profile Shape of D : Vindication of Using Ellipse Equation for $\ln(D)$...	57
3.6.2	Asymmetric Slip Profile	58
3.6.3	Similarity Between Different Stochastic Representations of Displacement.....	59
3.6.4	Probability Distribution of D	65
3.6.5	Range of Model Applicability	65
4	Example Hazard Applications.....	112
4.1	Numerical implementation for scenario events	112
4.2	Example applications	113
4.2.1	M 7.0 Scenario Example in Petersen et al. (2011).....	114
4.2.2	M 5.8 Scenario Example in IAEA Benchmarking Study	114
4.2.3	Hypothetical M 8.3 Scenario	115
4.2.4	Scenario Examples with Model Epistemic Uncertainty	116
4.2.5	Example in IAEA Benchmarking Study with Source Uncertainty.....	116
5	Summary and Future Work	126
5.1	Summary.....	126
5.2	Potential model improvements.....	127
5.3	Future model development.....	128
	REFERENCES	130
APPENDIX A	Petersen et al. (2011) PFDHA Framework.....	135
APPENDIX B	Data Selection Tables	142
APPENDIX C	Implementation of Probability Distributions.....	147

APPENDIX D Tier Classification of FDHI Principal Displacement Data.....	154
ELECTRONIC SUPPLEMENTS.....	168

LIST OF FIGURES

Figure 2.1. Geometric parameters used in fault displacement model development and hazard analysis (modified from Petersen et al., 2011). Variables l and Δ are utilized to specify the along-main-trace and off-main-trace position, respectively, of a point. Variable L is the length of the main trace. Variable s is utilized to track the position of main trace along the mapped fault trace.	20
Figure 2.2. Comparison of least-cost path (LCP) and event coordinate system (ECS) reference lines for the 1992 M 7.28 Landers earthquake. (A) View of the entire rupture length. (B) Close-up view of a stepover area.....	21
Figure 2.3. Comparison of measured displacement from Landers in the coordinate systems for least-cost path (LCP) and event coordinate system (ECS). (A) Along-strike distribution of principal displacement. (B) Strike-normal distribution of distributed displacement.	22
Figure 2.4. Aggregation of principal displacements on subparallel principal rupture traces. (A) Map view. (B) Individual and aggregated displacement.....	23
Figure 2.5. Segment and branch number assignment for Landers measurement points.....	24
Figure 2.6. Aggregation of principal displacement for segment 1 and branch 1.1 of Landers surface rupture. (A) Map view with measurement location color-coded by percent contribution to the aggregated displacement. (B) Individual and aggregated displacement.	25
Figure 2.7. Comparison of along-strike distributions of principal displacement. (A) As-measurement principal displacement. (B) Aggregated principal displacement.....	26
Figure 2.8. Epicenter distribution of strike-slip earthquakes in the Fault Displacement Hazard Initiative (FDHI) database (Sarmiento et al., 2021), including earthquakes selected for model development and those excluded for reasons discussed in Section 2.2.1.....	27
Figure 2.9. Distribution of displacement measurements plotted with magnitude (M) in along-strike distance bins. Each bin is defined by normalized location along the main trace (i.e., values of l/L , where L is rupture length and l is distance to the nearest end of rupture). Tier-1 and tier-2 data are explained in Section 3.2.2.1 and in Appendix D.....	28
Figure 2.10. Distribution of displacement measurements along normalized rupture location (l/L) for individual earthquakes. Tier-1 and tier-2 data are explained in Section 3.2.2.1 and in Appendix D.	29

Figure 2.11. Histogram of the number of slip measurements (N_{slip}) for an individual earthquake. Out of the 29 strike-slip earthquakes, four have more than 200 measurements, and all four are of M 7.1 or larger. These are the 1992 Landers ($N_{\text{slip}} = 566$), 1999 Izmit Kocaeli ($N_{\text{slip}} = 263$), 2013 Balochistan ($N_{\text{slip}} = 247$), and 2019 Ridgecrest mainshock ($N_{\text{slip}} = 226$)..... 30

Figure 3.1. Quantiles of fault displacement (D) predicted by the simple update. (A) Displacement plotted versus magnitude (M). For comparison, quantiles predicted by the original Petersen et al. (2011, P11) elliptical model are plotted as dashed lines. (B) Displacement plotted against the normalized position along the main trace ($l2L$). For comparison, quantiles predicted by P11 (Petersen et al., 2011) are plotted as dashed curves. 74

Figure 3.2. Residual diagnostic plots of the simple update to Petersen et al. (2011). Standardized residual ϵ/σ , where $\epsilon = \text{observed } \ln(D) - \text{predicted mean of } \ln(D)$ and $\sigma = \text{standard deviation of } \epsilon$, is used on this figure. (Top left) residuals versus $l2L_f$. Variable $l2L_f$ is the $l2L (= l/L)$ folded at the midpoint ($l/L = 0.5$) of the main trace, l is the along-main-trace position of the data point, and L is the length of the main trace. (Top right) residuals versus magnitude (M). (Middle left) histogram of residuals; probability density function (PDF) of the normal distribution fitted to the residuals is plotted as the red curve. (Middle right) empirical cumulative distribution function (CDF) of residuals; CDF of the normal distribution fitted to the residuals is plotted as the red curve. (Bottom left) quantile-quantile (Q-Q) plot that compares the quantiles of standardized residuals against the quantiles of standard normal distribution; the solid red line passes through the 1st and the 3rd quartiles, and the shaded area denotes roughly the 0.95 confidence level if residuals are taken from the standard normal distribution. 75

Figure 3.3. Data residuals of the simple update to Petersen et al., (2011), plotted versus magnitude (M). A residual is color coded according to the earthquake from which it is sampled. Solid triangle marks the mean of residuals in a particular earthquake ($\bar{\epsilon}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \epsilon_{ij}$, where n_i is the number of displacement measurements in earthquake i). The blue curve shows the result of a fixed-effect regression of the bilinear M -scaling function (function f_M of Equation (3.2)) to all data residuals. The red curve shows the result of a random-intercept mixed-effect regression, in which variation in the mean of individual earthquakes is modeled. The striking difference between these two fitted curves highlights the importance of proper weighting of individual earthquake afforded by the mixed-effect regression. 76

Figure 3.4. Within-earthquake residuals of the simple update to P11 (Petersen et al., 2011), plotted against $l2L_f$. Variable $l2L_f$ is the $l2L (= l/L)$ folded at the midpoint ($l/L = 0.5$) of the main trace, l is the along-main-trace position of the data point, and L is the length of the main trace. Within-earthquake residual is approximately computed as $(\epsilon_{ij} - \bar{\epsilon}_i)$

where $\bar{\varepsilon}_i$ is the mean of ε_{ij} in earthquake i , as defined in the text and in the caption of Figure 3.3. To visualize the variation of mean residual along the main trace, a loess smooth to all of the within-earthquake residuals is plotted as the red curve. A separate loess smooth to the tier-1 data residual is shown as the black curve. A comparison of these two loess smooths indicates that, relative to the tier-1 smooth, the estimated mean is decreased by an average of about 17%, due to the inclusion of tier-2 data. To assess its variation along the main trace, residual standard deviation (S.D.) in ten $l2L_f$ bins are computed and tag onto the loess smooths. Relative to tier 1's standard deviation, the standard deviation of residuals in an $l2L_f$ bin is increased by an average of about 27% along the main trace, due to the inclusion of tier-2 data. Tier-1 data's residual standard deviation reveals a slight downward trend with $l2L_f$. An assessment of the $l2L_f$ -dependence of residual standard deviation for teir-1 data is shown in Figure 3.5..... 77

Figure 3.5. Within-earthquake residuals of *Modell.NO*, plotted against $l2L_f$. Variable $l2L_f$ is the $l2L (= l/L)$ folded at the midpoint ($l/L = 0.5$) of the main trace, l is the along-main-trace position of the data point, and L is the length of the main trace. Solid red symbol marks the standard deviation of residuals in a particular $l2L_f$ bin multiplied by 1.65. Bin boundaries are shown as the vertical dotted lines. Exponential function $e^{a_0 + a_1 l2L_f}$ and linear functions $a_0 + a_1 l2L_f$ are fitted to the bin standard deivations. The fitted exponential and linear functions, mutiplied by 1.65, are shown as the blue and the red curves, respectively..... 78

Figure 3.6. Random intercept $\delta_{eq,i}$ of *Modell.NO*, plotted versus magnitude (**M**). Solid red circles mark the standard deviation of $\delta_{eq,i}$ in respective magnitude bins bounded by the vertical dotted red lines. The horizontal short-dashed line marks the estimated σ_{eq} assuming it is a constant. The fitted **M**-dependent σ_{eq} model (Euqation (3.4)) is written in the top left corner of the figure and shown as the red dashed curve. The horizontal long-dashed line indicates the imposed floor level of σ_{eq} in the large magnitude range..... 79

Figure 3.7. The 0.05, 0.50, and 0.95 quantiles (the 5th, 50th, and 95th percentiles in the figure title) of fault displacement (D) from the compound distribution of *Modell.NO*. (A) Quantiles versus magnitude **M**. (B) Quantiles versus the normalized position along the main ($l2L$). Quantiles predicted by P11 (Petersen et al., 2011) are shown as the long-dashed curves..... 80

Figure 3.8. Residual diagnostic plots of *Modell.NO*. Within-earthquake (EQ) standardized residuals are used on this figure. (Top left) residuals versus $l2L_f$. Variable $l2L_f$ is the $l2L (= l/L)$ folded at the midpoint ($l/L = 0.5$) of the main trace, l is the along-main-trace position of the data point, and L is the length of the main trace. (Top right) residuals versus magnitude (**M**). (Middle left) histogram of residuals; probability density function (PDF) of the normal distribution fitted to the residuals is plotted as the red

curve. (Middle right) empirical cumulative distribution function (CDF) of residuals; CDF of the normal distribution fitted to the residuals is plotted as the red curve. (Bottom left) quantile-quantile (Q-Q) plot that compares the quantiles of standardized residuals against the quantiles of standard normal distribution; the solid red line passes through the 1st and the 3rd quartiles, and the shaded area denotes roughly the 0.95 confidence level if residuals are taken from the standard normal distribution..... 81

Figure 3.9. Estimated random intercept $\delta_{eq,i}$ of *Model2.SN*, plotted versus magnitude (**M**). Solid red circles are the standard deviation of $\delta_{eq,i}$ in respective magnitude bins bounded by the vertical dotted red lines. The short-dashed line marks the estimated σ_{eq} assuming it is a constant. The fitted **M**-dependent σ_{eq} model is shown as the red dashed curve; the model is also written in the top left of this figure. The long-dashed line indicates the imposed floor level of σ_{eq} 82

Figure 3.10. Residual diagnostic plots of *Model2.SN*. Within-earthquake (EQ) normalized quantile residuals are used on this figure. (Top left) residuals versus $l2L_f$. Variable $l2L_f$ is the $l2L$ ($= l/L$) folded at the midpoint ($l/L = 0.5$) of the main trace, l is the along-main-trace position of the data point, and L is the length of the main trace. (Top right) residuals versus magnitude (**M**). (Middle left) histogram of residuals; probability density function (PDF) of the normal distribution fitted to the residuals is plotted as the red curve. (Middle right) empirical cumulative distribution function (CDF) of residuals; CDF of the normal distribution fitted to the residuals is plotted as the red curve. (Bottom left) quantile-quantile (Q-Q) plot that compares the quantiles of normalized quantile residuals against the quantiles of standard normal distribution; the solid red line passes through the 1st and the 3rd quartiles, and the shaded area denotes roughly the 0.95 confidence level if residuals are taken from the standard normal distribution. 83

Figure 3.11. The 0.05, 0.50, and 0.95 quantiles (the 5th, 50th, and 95th percentiles in the figure title) of fault displacement (D) from the compound distribution of *Model2.SN*. (A) Quantiles versus magnitude **M**. (B) Quantiles versus the normalized position along the main trace ($l2L$). Quantiles predicted by P11 (Petersen et al., 2011) are shown as the long-dashed curves..... 84

Figure 3.12. Estimated random intercept $\delta_{eq,i}$ of *Model3.ST*, plotted versus magnitude (**M**). Solid red circles are the standard deviations of $\delta_{eq,i}$ in respective magnitude bins bounded by the vertical dotted red lines. The short-dashed line marks the estimated σ_{eq} assuming it is a constant. The fitted **M**-dependent σ_{eq} model is shown as the red dashed curve; the model is also written in the top left of this figure. The long-dashed line indicates the imposed floor level of σ_{eq} 85

Figure 3.13. Residual diagnostic plots of *Model3.ST*. Within-earthquake (EQ) normalized quantile residuals are used on this figure. (Top left) residuals versus $l2L_f$. Variable $l2L_f$ is the $l2L$ ($= l/L$) folded at the midpoint ($l/L = 0.5$) of the main trace, l is the along-

main-trace position of the data point, and L is the length of the main trace. (Top right) residuals versus magnitude (**M**). (Middle left) histogram of residuals; probability density function (PDF) of the normal distribution fitted to the residuals is plotted as the red curve. (Middle right) empirical cumulative distribution function (CDF) of residuals; CDF of the normal distribution fitted to the residuals is plotted as the red curve. (Bottom left) quantile-quantile (Q-Q) plot that compares the quantiles of normalized quantile residuals against the quantiles of standard normal distribution; the solid red line passes through the 1st and the 3rd quartiles, and the shaded area denotes roughly the 0.95 confidence level if residuals are taken from the standard normal distribution. 86

Figure 3.14. The 0.05, 0.50, and 0.95 quantiles (the 5th, 50th, and 95th percentiles in the figure title) of fault displacement (D) from the compound distribution of *Model3.ST*. (A) Quantiles versus magnitude **M**. (B) Quantiles versus the normalized position along the main trace ($l2L$). Quantiles predicted by P11 (Petersen et al., 2011) are shown as the long-dashed curves..... 87

Figure 3.15. Estimated random intercept $\delta_{eq,i}$ of *Model4.nEMG*, plotted versus magnitude (**M**). Solid red circles are the standard deviations of $\delta_{eq,i}$ in respective magnitude bins bounded by the vertical dotted red lines. The short-dashed line marks the estimated σ_{eq} assuming it is a constant. The fitted **M**-dependent σ_{eq} model is shown as the red dashed curve; the model is also written in the top left of this figure. The long-dashed line indicates the imposed floor level of σ_{eq} 88

Figure 3.16. Residual diagnostic plots of *Model4.nEMG*. Within-earthquake (EQ) normalized quantile residuals are used on this figure. (Top left) residuals versus $l2L_f$. (Top right) residuals versus magnitude **M**. (Middle left) histogram of residuals; probability density function (PDF) of the normal distribution fitted to the residuals is plotted as the red curve. (Middle right) empirical cumulative distribution function (CDF) of residuals; CDF of the normal distribution fitted to the residuals is plotted as the red curve. (Bottom left) quantile-quantile (Q-Q) plot that compares the quantiles of normalized quantile residuals against the quantiles of standard normal distribution; the solid red line passes through the 1st and the 3rd quartiles, and the shaded area denotes roughly the 0.95 confidence level if residuals are taken from the standard normal distribution. 89

Figure 3.17. The 0.05, 0.50, and 0.95 quantiles (the 5th, 50th, and 95th percentiles in the figure title) of fault displacement (D) from the compound distribution of *Model4.nEMG*. (A) Quantiles versus magnitude **M**. (B) Quantiles versus normalized position along the main trace ($l2L$). For comparison, quantiles predicted by P11 (Petersen et al., 2011) are shown as the long-dashed curves. 90

Figure 3.18. Comparison of compound distributions of fault displacement (D) from the models summarized in Table 3.1 and from Petersen et al. (2011) (P11). (Left)

probability density functions (PDFs). (Right) cumulative distribution functions (CDFs)
 Calculations are carried out for ruptures of four different magnitudes (**M**), all at the
 same normalized position of $l2L = 0.2$ 91

Figure 3.19. Comparison of the 0.05 (Q05), 0.50 (Q50), and 0.95 (Q95) quantiles of fault displacement (D) from the compound distributions of models summarized in Table 3.1 and from the Petersen et al. (2011) (P11). The rupture magnitude (**M**) and the site's normalized position ($l2L$) are indicated in the panel title. 92

Figure 3.20. Quantiles of compound distribution of fault displacement (D) from the models summarized in Table 3.1 and from Petersen et al. (2011) (P11), plotted against magnitude (**M**). Each panel is conditional on the quantile ($Q05$, $Q50$, and $Q95$ for the 0.05-quantile, 0.5-quantile, and 0.95-quantile, respectively) and the normalized position along the main trace ($l2L = 0, 0.2, \text{ and } 0.5$) as indicated in the strip of each panel. 94

Figure 3.21. Magnitude (**M**)-scaling relations of fault displacement models (FDMs) regressed using an m_3 fixed to one of the five different values indicated in the plot legend. These m_3 values are marked in this figure by the color-coded vertical lines. For use as a reference for model comparison, the earthquake term ($c_0 + c_{eq,i}$) obtained from a regression analysis of an exploratory FDM without the magnitude scaling term f_M are plotted as the solid squares. Note that the five FDMs shown on this figure are based on the $\ln(D_{ij})$ data, not on the $(c_0 + c_{eq,i})$ values from the exploratory FDM..... 95

Figure 3.22. Quantiles of compound distributions of fault displacement (D) from the four models listed in Table 3.6, plotted versus magnitude (**M**). The fixed m_3 value used in each model is indicated in the figure legend. For reference, the quantiles from the model assuming $m_3 = 7.45$ are shown as the gray dashed curves. Each panel is conditional on the quantile ($Q05$, $Q50$, and $Q95$ for the 0.05-quantile, 0.5-quantile, and 0.95-quantile, respectively) and the normalized position along the main trace ($l2L = 0, 0.2, \text{ and } 0.5$) as indicated in the strip of each panel. 96

Figure 3.23. Profile of median slip predicted by *Model4.nEMG* for magnitude (**M**) 7.5 The predicted median is fitted by three functions of $l2L$ used in previous studies to characterize slip profile along the main trace. The fitted curves are shown as dashed curves in red, magenta, and orange color for the symmetric isosceles trapezoid, hyperbolic tangent (*tanh*) function, and ellipse equation, respectively. 97

Figure 3.24. Plots of normalized quantile residuals from *Model4.nEMG*. (Top) The horizontal coordinate is the original normalized distance $l2L$ measured relative to the western end point of main trace. (Bottom) The horizontal coordinate is the normalized distance $l2L$ after left-skewed profiles are reflected. A linear function of $l2L$ is fitted to the residuals in each panel and the fitted curves are shown as the dashed lines..... 98

Figure 3.25. Predicted quantiles of displacement (D) versus normalized position along the main trace ($l2L$), for a magnitude (M) 7.5 rupture. Black curves are computed from *Model4.nEMG*, a spatially symmetric model. Red curves are computed from the spatially asymmetric *Model6.nEMG*. Note that the peak of predicted quantile from *Model6.nEMG* occurs near $l2L = 0.4$ 99

Figure 3.26. The average (D_{ave}) and the maximum (D_{max}) of observed displacements for each strike-slip earthquake selected from the Fault Displacement Hazard Initiative (FDHI) database, plotted against magnitude (M). The fits by the bilinear function of M (Equations (3.23) and (3.24)) are shown as the red curve and the black curve for D_{ave} and D_{max} , respectively. For reference, the M -scaling relation $c_0 + f_M$ of *Model4.nEMG* is shown as the blue curve..... 100

Figure 3.27. Logarithm of the ratio D_{ave}/D_{max} , plotted against magnitude (M). The average (D_{ave}) and the maximum (D_{max}) of observed displacements in an individual earthquake are obtained from two data sources. The blue circles are from the Fault Displacement Hazard Initiative (FDHI) database (Sarmiento et al., 2021), and the red circles are from Wells and Coppersmith (1994). The average value of $\ln(D_{ave}/D_{max})$ from each dataset is marked by the horizontal dashed line..... 101

Figure 3.28. (left) Logarithm of observed maximum displacement ($\ln(D_{max})$) versus the magnitude-scaling term ($\delta_{eq,i} + c_0 + f_M$) of *Model4.nEMG*. (right) Logarithm of observed average displacement ($\ln(D_{ave})$) versus ($\delta_{eq,i} + c_0 + f_M$) of *Model4.nEMG*. Both $\ln(D_{max})$ and $\ln(D_{ave})$ have a 1:1 relation with the ($\delta_{eq,i} + c_0 + f_M$) term. The standard deviation of the scatter around the 1:1 line is 0.2548 and 0.2353 for $\ln(D_{max})$ and $\ln(D_{ave})$, respectively..... 102

Figure 3.29. Predicted probability density function (PDF) of fault displacement (D) versus the normalized position along the main trace ($l2L$) of a magnitude (M) 7.2 rupture. The predicted PDF conditional on zero random intercept ($\delta_{eq,i} = 0$) is shown by the gray curve. The predicted mean of the Gaussian component (μ) (magenta square) falls between the predicted 90th and 95th percentiles of displacements. The predicted mean displacement ($\mu - v$, where v is the mean of the Exponential component) (blue square) falls below the 50th percentile, as expected for a left-skewed distribution. For reference, the PDF of the compound distribution resulting from the marginalization over δ_{eq} is shown as the red curve. 103

Figure 3.30. Simulated exceedence probability of sample maximum, plotted against sample size. Samples are taken from the predicted nEMG distribution for magnitude (M) 7.3 at the normalized along-trace location of $l2L = 0.5$, for eight different sample sizes. To account for sampling variability, sampling is repeated 100 times for each sample size. The red solid square marks the mean of exceedence probability over these 100 trials. (Inset) Histogram of the sample size of displacements, per earthquake, in Fault

Displacement Hazard Initiative (FDHI) database that fall within a 1-km-wide spatial window centered at the location of observed maximum displacement (D_{max})..... 104

Figure 3.31. Matching the logarithm of Weibull-distributed ratio of displacement (D) to average displacement (D_{ave}), D/D_{ave} , by the negative exponentially modified Gaussian (nEMG) distribution. (Left) Histogram of the logarithm of D/D_{ave} sampled from a Weibull distribution whose parameters, as shown in the panel title, are predicted by the D/D_{ave} model of Moss and Ross (2011; MR11). The density function of the matched nEMG distribution is shown as the red curve. Parameters of the matched nEMG distribution are given in the plot legend. (Right) Quantile-quantile plot that compares quantiles of the logarithm of sampled D/D_{ave} versus the theoretical quantile of the matched nEMG distribution..... 105

Figure 3.32. Matching the logarithm of gamma-distributed ratio of displacement (D) to average displacement (D_{ave}), D/D_{ave} , by the negative exponentially modified Gaussian (nEMG) distribution. (Left) Histogram of the logarithm of D/D_{ave} sampled from a gamma distribution whose parameters, as shown in the panel title, are predicted by the D/D_{ave} model of Moss and Ross (2011; MR11). The density function of the matched nEMG distribution is shown as the red curve. Parameters of the matched nEMG distribution are given in the plot legend. (Right) Quantile-quantile plot that compares quantiles of the logarithm of sampled D/D_{ave} versus the theoretical quantile of the matched nEMG distribution..... 106

Figure 3.33. Matching the logarithm of beta-distributed ratio of displacement (D) to average displacement (D_{max}), D/D_{max} , by the negative exponentially modified Gaussian (nEMG) distribution. (Left) Histogram of the logarithm of D/D_{max} sampled from a beta distribution whose parameters, as shown in the panel title, are predicted by the D/D_{max} model of Moss and Ross (2011; MR11). The density function of the matched nEMG distribution is shown as the red curve. Parameters of the matched nEMG distribution are given in the plot legend. (Right) Quantile-quantile plot that compares quantiles of the logarithm of sampled D/D_{max} versus the theoretical quantile of the matched nEMG distribution. 107

Figure 3.34. Matching the logarithm of gamma-distributed ratio of displacement (D) to maximum displacement (D_{max}), D/D_{max} by the negative exponentially modified Gaussian (nEMG) distribution. (Left) Histogram of the logarithm of D/D_{max} sampled from a gamma distribution whose parameters, as shown in the panel title, are predicted by the D/D_{max} model of Moss et al. (2022; MEA22). The density function of the matched nEMG distribution is shown as the red curve. Parameters of the matched nEMG distribution are given in the plot legend. (Right) Quantile-quantile plot that compares quantiles of the logarithm of sampled D/D_{max} versus the theoretical quantile of the matched nEMG distribution. 108

- Figure 3.35.** Scatter plots of observed average displacement (D_{ave}) versus magnitude (M). (Top left) data from Wells and Coppersmith (1994; WC94); (Top right) data from Wells and Youngs (2015; WY15); all styles of faulting (SOF); (Bottom left) data from Anderson et al. (2021; AEL21); (Bottom right) Fault Displacement Hazard Initiative (FDHI) database (Sarmiento et al., 2021) used in this study. A *loess* smooth is calculated and plotted as solid black curve in each panel. For reference, three parametric models fitted to the datasets of WC94, FDHI, and AEL21 are shown as dashed curves in black, red, and blue color, respectively..... 109
- Figure 3.36.** Data residuals with respect to the D_{ave} - magnitude (M) relation of Equation (3.23) developed using the average displacement (D_{ave}) data from the Fault Displacement Hazard Initiative (FDHI) database (Sarmiento et al., 2021). To help visualize the residual trend, the smooth curve fitted by the nonparametric, locally weighted (*loess*) regression on the data residuals is shown as the dashed curve..... 110
- Figure 3.37.** The magnitude-scaling terms ($c_0 + f_M$) from *Model4.nEMG* and its three epistemic variants, all vertically shifted by -0.7642 log units, are plotted against magnitude (M). For comparison, average displacement (D_{ave}) data from four different data sources (Wells and Coppersmith (1994; WC94); Wells and Youngs (2015; WY15); Anderson et al. (2021; AEL21) and Fault Displacement Hazard Initiative (FDHI) database (Sarmiento et al., 2021)) are plotted in four separate panels as the solid gray symbols. WY15 data include strike-slip as well as other styles of faulting (SOF)..... 111
- Figure 4.1.** Comparison of hazard curves for the magnitude (M) 7.0 scenario earthquake using the preferred negative exponentially modified Gaussian (nEMG) model (*Model4.nEMG*), the Petersen et al. (2011) bilinear model (P11 Bilinear), and two interim models (*Model1.NO* and *Model2.SN*). $P[sr \neq 0|m]$ is the probability of having surface rupture (i.e., $sr \neq 0$) given that a magnitude m earthquake occurs. 119
- Figure 4.2.** Comparison of hazard curves for the M 5.8 scenario earthquake using the preferred negative exponentially modified Gaussian (nEMG) model, the Petersen et al. (2011) bilinear model (P11 Bilinear), and two interim models (*Model1.NO* and *Model2.SN*). 120
- Figure 4.3.** Comparison of hazard curves for the hypothetical magnitude (M) 8.3 scenario using the preferred negative exponentially modified Gaussian (nEMG) model, the Petersen et al. (2011) bilinear model (P11 Bilinear), and two interim models (*Model1.NO* and *Model2.SN*). 121
- Figure 4.4.** Hazard curves showing effects of alternative m_3 values in models using negative exponentially modified Gaussian (nEMG) distribution. (A) Magnitude (M) 7.0 scenario in Petersen et al. (2011). (B) M 5.8 scenario example in International Atomic Energy Agency (IAEA) benchmarking study (Valentini et al., 2021). (C) Hypothetical M 8.3 scenario. 122

Figure 4.5. Hazard curves showing effects of alternative assumptions for along-strike slip distribution using the magnitude (**M**) 7.0 scenario as an example. (A) $l/L = 0.5$. (B) $l/L = 0.4$. (C) $l/L = 0.3$. (D) $l/L = 0.1$. Variable l is the along-main-trace position and L is the length of the main trace (see Figure 2.1). 123

Figure 4.6. Example in the International Atomic Energy Agency (IAEA) benchmarking study with source model logic tree (modified from Valentini et al., 2021). (A) Three segments of the Fudagawa fault zone. (B) Rupture scenarios. (C) Source logic tree with epistemic uncertainty in magnitude and event rate. M is earthquake magnitude; l is the along-main-trace position, and L is the length of the main trace (see Figure 2.1). 124

Figure 4.7. Comparison of Petersen et al. (2011; P11) bilinear and negative exponentially modified Gaussian (nEMG) hazard curves for a site on the Fudagawa fault zone in Japan. Calculated mean hazard curve is the sum of weighted mean of all scenario curves. (A) Mean and percentile hazard curves. (B) Scenario and total hazard curves. M is earthquake magnitude. 125

Figure A.1. Illustration of the Petersen et al. (2011) and Chen and Petersen (2019) probabilistic fault displacement hazard analysis framework, component models, and data needed to develop empirical formula of each component model. 141

Figure D.1. An example of tier classification of the Landers principal net displacement (D , in units of meters) from the preferred data source PT_DS_ID = 6, using the 2-parameter ellipse functional form. The estimated 0.997 and 0.5 quantiles are shown as the orange and red curves, respectively. The estimated 0.003 quantile of the hypothetical normal distribution is shown as the thick red curve, while the lower-bound of tier-1 displacements is shown as the thick black curve. The identified tier-2 displacements are marked by a small solid square inside an open square. Displacements from the same rupture segment are marked by the same color. For comparisons, the 0.15 and the 0.1 quantiles are shown as the cyan and the blue dashed lines, respectively. 158

Figure D.2. Tier classification of principal net displacement (D) obtained using each of the eight alternative functional form of $f_{l2L}(l2L)$: (A) Elliptical. (B) Elliptical plus Linear. (C) Quadratic. (D) Cubic. (E) Modified Beta. (F) B-Spline, degrees of freedom (df) = 3. (G) B-Spline df = 5. (H) B-Spline df = 7. Earthquake name, magnitude (M), and preferred FDHI data source identification (DS) are indicated inside each panel title strip. 159

Figure D.3. Single final tier classification of a data point. Earthquake name, magnitude (M), and preferred FDHI data source identification (DS) are indicated inside the panel title strip. Indicator variable $flag_i$ is equal to 1 if the data point of interest is classified as tier 2 by the i -th functional form, otherwise $flag_i = 0$. The sum ($\sum_{i=1}^8 flag_i$) is the number of times a data point is classified as tier 2. A data point is assigned a final classification of tier-2 if ($\sum_{i=1}^8 flag_i$) > $recom$; that is, it is classified as tier 2 more than $recom$ times. The criterion $recom$ is earthquake specific and given inside each panel. 167

LIST OF TABLES

Table 3.1. Coefficients of Interim Updates and Preferred Model.....	68
Table 3.2. Distribution Parameters and Percentiles Predicted by <i>Model1.NO</i>	69
Table 3.3. Distribution Parameters and Percentiles Predicted by <i>Model2.SN</i>	70
Table 3.4. Distribution Parameters and Percentiles Predicted by <i>Model3.ST</i>	71
Table 3.5. Distribution Parameters and Percentiles Predicted by <i>Model4.nEMG</i>	72
Table 3.6. Coefficients of nEMG Distributional Models, Conditional on Different Credible m_3	73
Table B.1. Strike-slip events in the Fault Displacement Hazard Initiative (FDHI) database ...	143
Table B.2. Rupture dataset selection and rationale	144
Table B.3. Explanation of data quality indices in the Fault Displacement Hazard Initiative (FDHI) database (October 2020 release) and usage recommendation (slightly simplified) ¹	145
Table B.4. Measurement dataset selection and rationale	146

1 Introduction

Probabilistic fault displacement hazard analysis (PFDHA) is a quantitative assessment of hazards associated with potential surface ruptures from future earthquakes. It was first developed by Youngs et al. (2003) to assess hazards from the numerous normal active faults that affect the site of the proposed repository for high-level nuclear wastes at Yucca Mountain, Nevada. They developed two PFDHA approaches: an earthquake approach and a displacement approach. They also established probability distributions appropriate for normal faults in an extensional tectonic environment.

The earthquake approach for PFDHA is similar to the well-established probabilistic seismic hazard analysis (PSHA) for ground motion hazards. It is often viewed as PSHA with ground motion models replaced by fault displacement models (FDMs) that characterize how the amplitude of fault displacement varies along an earthquake's rupture trace and how it attenuates away from the surface rupture. The earthquake approach, therefore, can take advantage of the best available scientific models in source characterization developed for PSHA. PFDHA, however, demands much higher precision in fault location and needs to take into consideration greater details in fault geometric complexity (Petersen et al., 2011, hereafter P11; Chen and Petersen, 2011; Chen et al., 2013). Another important aspect of PFDHA that distinguishes it from PSHA is that PFDHA requires additional probability models to account for the probability of occurrence of surface ruptures because not all earthquakes produce surface rupture and, for earthquakes that do produce surface ruptures, surface ruptures do not occur everywhere near or along the earthquake fault.

The displacement approach relies on establishing statistical characteristics of site-specific displacements and models for recurrence of displacement events. It treats each surface rupture event as an independent event without reference to the causative earthquake. It is computationally and conceptually simpler than the earthquake approach but requires sufficient site-specific displacement measurements and occurrence times to establish statistical models.

Since the work of Youngs et al. (2003), the PFDHA methodology has been improved and its applicability extended to broader tectonic environments through the work of P11, Moss and Ross (2011), Takao et al. (2013), and Nurminen et al. (2020). The P11 work is a collaborative effort between the U.S. Geological Survey (USGS) and California Geological Survey (CGS) that extended the Youngs et al. (2003) approach to strike-slip faults and to account for fault trace mapping accuracy, fault complexity, and footprint size of building/engineered structures. The P11

PFDHA framework and hazard equations are summarized in Appendix A of this report. Using the P11 approach, advances were made to create fault displacement hazard maps (Figure 11 in P11; and Figure 8 in Chen and Petersen, 2011). These maps are similar to seismic hazard maps for ground motion, such as seismic hazard maps in Petersen et al. (2020). Fault displacement hazard maps may supplement the regulatory Alquist-Priolo Earthquake Fault Zone maps (<https://www.conservation.ca.gov/cgs/alquist-priolo>) produced by CGS for more efficient mitigation of risk associated with surface displacement hazards where avoidance of active faults is not feasible, such as with pipelines and transportation corridors. They may also supplement ground motion hazard maps for quantifying overall hazards from earthquakes in the vicinity of active faults.

In recent years, PFDHA has been used increasingly to assess displacement hazards for lifeline systems, nuclear power plants, and other critical structures (AMEC Geomatrix, 2010; U.S. Nuclear Regulatory Commission, 2012; Takao et al., 2013; Japan Nuclear Safety Institute, 2013; Rizzo, 2013a, b). Comprehensive assessment of fault displacement hazards is particularly important for the resilience and recovery of large population centers in California (such as Los Angeles metropolitan area and San Francisco Bay Area) and similar places around the world, where water, natural gas, electricity, and other vital supplies depend on engineered systems that cross large active faults. The American Nuclear Society (2015) published American National Standard Criteria for assessing tectonic surface fault rupture and deformation at nuclear facilities. The International Atomic Energy Agency (IAEA) requires its Member States to conduct probabilistic evaluations of surface faulting for nuclear installations (IAEA, 2016). The IAEA recently published a technical document (TECDOC) reviewing the state of practice in PFDHA (IAEA, 2021) and is in the process of preparing a second TECDOC providing detailed guidance to its Member States on PFDHA through a benchmarking study (Valentini et al., 2021).

However, PFDHA still relies on the few aforementioned published FDMs: Youngs et al. (2003) for normal faults, P11 for strike-slip faults, and Moss and Ross (2011) for reverse faults. These FDMs were developed using older datasets that are typically sparsely populated along the rupture trace and from a limited number of surface rupturing earthquakes prior to year 2000.

Recognizing the need to develop surface rupture and fault displacement data in a consistent manner to improve empirical models and advance PFDHA, an international effort was initiated to construct a worldwide and unified surface displacement database, the SURE (SURface Rupture of Earthquakes) database (Baize et al., 2020).

Started in 2018, a comprehensive database building effort was initiated as part of the fault displacement hazard initiative (FDHI) project led by Professor Yousef Bozorgnia at University of California, Los Angeles (Bozorgnia et al., 2021). The FDHI database includes global surface rupturing historical earthquakes of all faulting styles (Sarmiento et al., 2021). It contains displacement measurements whose observation locations are defined by geographic coordinates. It also contains georeferenced surface rupture traces, defined by geographic coordinates of vertices, and also given as polyline features in ArcGIS shapefile format. Another important

component of the FDHI research project is a collaborative community effort to develop new FDMs by multiple modeling teams. The authors of this report constitute one of the teams participating in this effort.

Our overall goal is to use the new FDHI database to improve empirical FDMs in P11. An overarching objective is to improve the FDMs for principal displacement (i.e., displacement associated with principal ruptures).

This report documents teamwork carried out since 2019 to improve the FDMs for principal displacement. It consists of three main chapters. Chapter 2 describes selection of principal displacement data, establishment of a reference coordinate system, and data preparation. Chapter 3 presents model development, including choices of functional forms, regression analysis methods and results, and model evaluation and comparison. Chapter 4 presents the outcomes of example applications of the improved FDMs in the PFDHA framework of P11.

As described in Appendix A, FDMs constitute one component of the P11 PFDHA framework. They are needed in the calculation of the conditional exceedance probability of displacement, i.e., the $P[D \geq D_0 | L, m, D \neq 0]$ term in Figure A.1 and in Equation (A.1). Other components in the P11 framework are not updated. These include distributed displacement model, models for surface rupture probability, and location uncertainty. The FDHI database can be used to update some of these models in the future. It does not contain data needed to define conditional probability of surface rupture given magnitude.

Our analysis is limited to fault displacement measured on discrete rupture from strike-slip events. Continuous deformation between discrete ruptures is not accounted for. We note that discrete displacement accounts for only a portion of the total deformation across earthquake faults. A substantial amount of continuous deformation can occur (Milliner et al. 2015, 2016) and can be damaging to structures built across the deformation zone. A parallel study led by Milliner et al. (2020) is ongoing. That study processes and develops geodetic data, establishes component models, and formulates methods suitable for hazard analysis of total displacement.

2 Rupture and Displacement Data

We selected mapped surface ruptures and displacement measurement data from the October 2020 release of the FDHI database (Sarmiento et al., 2021), supplemented by three strike-slip events added to the FDHI database in July 2022. Data selection is discussed in this chapter. Selected data are documented in the electronic supplements. This chapter also discusses the development of a piecewise continuous main rupture trace as the along-strike reference trace (the main trace), with respect to which the coordinate system is defined and principal displacements across multiple subparallel principal rupture traces are aggregated. Developing a reference trace and aggregating displacement are two challenging, yet necessary steps in data preparation for model development. These data processing procedures resulted in several additional attribute fields, which are included and described in the electronic supplements.

2.1 FDHI DATABASE

The FDHI database includes two main data tables: high resolution surface rupture traces and displacement values at measurement points. Both rupture traces and displacement measurement locations are georeferenced. Surface ruptures are defined by geographic coordinates of vertices in a flatfile (in comma-separated values file format) and as polyline features in the ArcGIS shapefile format. For some events, data are from multiple sources for displacement measurements and/or rupture traces. In this report, we refer data from a specific source as a dataset. Multiple measurement datasets and rupture datasets are available for some events.

In addition to geographic coordinates, the FDHI database also provides strike-parallel (u) and strike-normal (t) coordinates for measurement points and rupture trace vertices in an event-specific, local coordinate system (ECS) in which the ECS is the second generalized coordinate system (GC2) (Spudich and Chiou, 2015) relative to a smooth along-strike reference curve. The ECS reference curve is developed as a geometry center line of all displacement measurement points and rupture trace vertices, with displacement points weighted higher in accordance with the value of measured net displacement. Details of the ECS can be found in Section 4.5 of the database report (Sarmiento et al., 2021). ECS is used as the along-strike reference trace by most FDHI

model development teams. Our team used a least-cost path (LCP) analysis method (see Section 2.3) to determine a main rupture trace from mapped surface ruptures and use it as the reference trace for reasons summarized in Section 2.3.4.

The October 2020 release of the FDHI database contains rupture and displacement data for 66 global earthquakes. Among them, 31 are strike-slip events. Four strike-slip events were added to the FDHI database as part of its July 2022 update. The update brings the total number of strike-slip events to 35. The geographic distribution of epicenters of all original FDHI events are shown on Figure 1.1 of the database report (Sarmiento et al. 2021).

Displacement measurements in the FDHI database are given as vector components as well as net displacement at each measurement location. Net displacement is the vector sum of fault parallel slip, fault-normal slip, and vertical slip or scarp height. The FDHI project recommends model development teams to analyze and develop FDMs for net slip. Therefore, in addition to limiting our current study to strike-slip events, we also limit the displacement parameter to the net slip given in the FDHI database. See Figure 2.10 in Sarmiento et al. (2021) for definitions of slip components.

In published PFDHA frameworks (e.g., Youngs et al., 2003; P11), surface ruptures and displacements were classified into principal and distributed categories, and the two categories were analyzed separately. Following these studies and per request from model development teams, the FDHI database team classified each mapped surface rupture into either principal or distributed categories based on geological interpretations and following the definitions given in Youngs et al. (2003) and in P11. Specific criteria are discussed in Section 4.3 of the database report (Sarmiento et al., 2021). Measurement points are associated with surface ruptures and, for measurements on individual discrete ruptures (i.e., offset of piecing points on two sides of a rupture, see Figure 2.12 in Sarmiento et al., 2021), they are also classified into either principal or distributed categories. However, some measurements in the FDHI database are not on individual ruptures; they reflect displacement accumulated over an aperture of some finite width or over the entire deformation zone. Therefore, two additional categories are included in displacement measurements. They are the total displacement and cumulative displacement. Total displacement is differential displacement across the entire rupture zone, usually calculated from geodetic measurements. Cumulative displacement is displacement summed across a limited width of the rupture zone across the main rupture. Both total and cumulative displacement may include principal displacement, distributed displacement, and continuous deformation in between discrete ruptures. Detailed description for each is given in Section 4.3 and Table 4.2 of the database report (Sarmiento et al., 2021).

We decided to sum up principal net displacements across multiple subparallel principal ruptures for use in FDM as the response variable. The method used to aggregate principal displacement is discussed in Section 2.5 of this report. Most other FDHI model development teams also aggregate displacement. However, aggregation methods vary, and some teams aggregate not only principal displacement but also distributed displacement (e.g., Kuehn et al., 2022).

Locations of mapped ruptures in the ArcGIS polyline format are used to construct an along-strike reference trace for each event, which we refer to as the main rupture trace. The construction of the main trace using the least-cost-path analysis is described in Section 2.3. The main rupture trace has two main utilities in our data preparation:

- It is utilized as the reference trace in the calculation of GC2 coordinate for each measurement point. The u -coordinate is treated as the along-main-trace distance l for FDM development. We prefer this coordinate over the u -coordinate of ECS for reasons discussed in Section 2.3.3 and 2.3.4.
- It is utilized as the reference trace in aggregating the principal net displacement across multiple subparallel principal ruptures (Section 2.4).

2.2 DATA SELECTION

As stated earlier, we focus our study on strike-slip events and exclude all events with other faulting styles, including normal-oblique and reverse-oblique. There are two main motivations for this decision: (1) our main objective is to improve the displacement models in P11 that were developed for strike-slip events; and (2) previous studies have focused on events with a specific faulting style [i.e., Youngs et al. (2003) on normal faulting events, Moss and Ross (2011) on reverse faulting events, and P11 on strike-slip events] because events with different faulting styles behave differently. Normal-oblique events are excluded because whether they behave more like normal events or more like strike-slip events is unclear. For the same reason, reverse-oblique events are excluded.

Because we limit our analysis to net displacement derived from slip components measured on individual ruptures, we do not analyze displacement measurements classified as *total* or *cumulative* in the FDHI database except for a couple of events as discussed in Section 2.2.1. We further restrict our analysis to principal displacement only, because principal and distributed displacement have distinctly different characteristics.

Not all principal displacement data from strike-slip events in the FDHI database are used in our statistical analysis for various reasons. In some cases, data from the entire event are excluded. In other cases, certain measurements in the selected events are excluded. Reasons for exclusion vary from event to event, but in general, affected events are those with the following: (1) no principal displacement measurements; (2) the spatial extent of rupture mapping is incomplete relative to the spatial extent of the known surface rupture; (3) ruptures and measurements from aftershocks; or (4) few good quality displacement measurements or measurements that have poor spatial coverage. Also, we do not mix data from different sources for a given event. Therefore, for events with multiple rupture or measurement datasets, only one rupture data source and one measurement data source are selected.

2.2.1 Excluded Events and Rationale for Exclusion

The 35 strike-slip events in the FDHI database are listed in Table B.1 of Appendix B, in which earthquake names are those given as an attribute in the FDHI database. These names are used throughout this report. We excluded six events in our data analysis for model development: GalwayLake, HomesteadValley, Hualien, IzuOshima, IzuPeninsula, and Pisayambo. The remaining 29 events occurred between 1953 and 2019, with magnitude ranging from 6.00 to 7.90. Thirteen of the 29 events ($\approx 45\%$) occurred in California.

Darfield and Imperial1940 were excluded initially because all their displacement measurements are classified as cumulative. The FDHI database team indicated that the cumulative displacement for these events is comparable to the aggregated principal displacement. For Imperial1940, displacement measurements are classified as cumulative because principal slips on individual subparallel principal ruptures were added. For Darfield, the cumulative slip is also mainly from principal ruptures even though it may contain warping within about 20 m (Alexandra Sarmiento, UCLA, email communication, 2022). Because these data appear comparable with aggregated principal displacement that we model (see Section 2.4), we decided to bring these two events back to augment usable data.

In the October 2020 database that we used initially, Imperial1940 was poorly sampled. In May 2021, the database team added the along-strike displacement data profile established by Rockwell and Klinger (2013) to supplement the data in the original FDHI database. However, the Rockwell and Klinger (2013) dataset only covers a small portion of the Imperial1940 surface rupture. Also, this dataset was sampled at much higher frequency. It shows much greater along-strike variability and, overall, appears incomparable with displacement measurements from the rest of the Imperial1940 surface rupture or with data from other strike-slip events. We decided to use the dataset in the original October 2020 database and not the Rockwell and Klinger (2013) dataset as the preferred data source for Imperial1940.

GalwayLake has only 8 usable displacement measurements. Almost all these measurements are on the southern part of the surface rupture. HomesteadValley has only 4 usable principal displacement measurements.

For Hualien, available data only cover a small portion of the surface rupture onshore of Taiwan. Most of its rupture is offshore to the northeast, under the Pacific Ocean. Unmapped surface rupture also is likely to the southwest.

Surface rupture and measurement data for IzuOshima and IzuPeninsula are only on Honshu Island. Surface ruptures offshore to the southeast are undocumented.

Pisayambo has only 11 displacement measurements. These measurements have amplitudes that are much higher than expected given the magnitude. Also, rupture mapping is at a much lower resolution than other events. The entire surface rupture is mapped as a single continuous line.

2.2.2 Dataset Selection

We decided to allow only one selected data source for events that have multiple datasets to avoid the need to (1) develop and justify a method to merge different sources of data, and (2) track the effects of individual datasets on regression analysis and results. Justifying selection of a particular dataset is easier and more straightforward. The decision to work with a selected dataset was also based on the following considerations: (1) rupture traces and displacement measurements that complement each other should be dealt with differently from those that supplement each other; however, determining which measurement points or rupture traces from different datasets supplement each other and which complement each other is difficult and time consuming; (2) different datasets often have different mapping resolution and different measurement accuracy so mixing them would not be ideal. We also preferred not to salvage data entries from datasets that are not selected because data salvation is complicated to perform, its outcome is difficult to justify, and benefits are not obvious.

In the early stage of our study, one exception to our one dataset per event rule is Balochistan. This event exerts large influences on the estimated magnitude scaling, particularly in the October 2020 version of the FDHI database. We initially analyzed principal displacement data from the two Balochistan datasets separately. After adding three strike-slip events from the July 2022 database update, the influence of Balochistan is diluted, and the single-dataset rule was again enforced.

Among the 29 selected events listed in Table B.1, four have two rupture datasets: Darfield, Duzce, Ridgecrest1, and Ridgecrest2. Rupture dataset selection is mainly based on FDHI database team’s recommendation (a pdf file titled “Notes on alternative rupture datasets,” distributed as part of the October 2020 FDHI database release). In addition, we consider completeness of spatial coverage because it affects the estimation of rupture length, which is a critical parameter in data analysis and model development. We also prefer the same mapped rupture data source as the preferred point measurement dataset (i.e., companion rupture and measurement datasets). Table B.2 summarizes the selected and excluded rupture datasets and primary reasons for dataset selection. FDHI database duplicates each entry in the flatfile and shapefile for measurement data when there are two rupture datasets. We eliminated duplicated entries associated with excluded rupture datasets.

Selection of measurement dataset is largely based on availability of high-quality measurements that are classified as principal. For each measurement point, the FDHI database offers two attributes (*net_flag* and *qual_code*) to indicate data quality and suggested usage. Three categories are in *net_flag*, and multiple codes are in *qual_code*. Explanation of these indices from the FDHI database report are reproduced in Table B.3 for easy reference.

Among the 29 selected events listed in Table B.1, 14 have multiple measurement datasets. Criteria for measurement dataset selection include (1) dataset that matches the selected rupture dataset (i.e., companion rupture and measurement datasets by the same authors); (2) dataset with

the largest number of high-quality principal displacement measurements; and (3) dataset that has more complete spatial coverage.

A measurement dataset that matches the rupture dataset is more likely to have measurement points that are located on mapped ruptures, avoiding potential error or inaccuracy induced from projecting measurement point to the nearest rupture. Companion datasets also mean both measurements and rupture maps would have similar accuracy and detail levels of study. A larger number of high-quality measurements is beneficial in regression analysis for model development. Table B.4 summarizes selected and excluded measurement datasets and primary reasons for dataset selection.

2.2.3 Excluded Ruptures and Measurements in Selected Datasets

In the FDHI database, each mapped rupture polyline in a given dataset is assigned a unique rupture identification number (*RUP_ID*). The significance of *RUP_ID*, however, is not obvious. Often a continuous rupture trace consists of multiple rupture lines (i.e., multiple *RUP_ID*s).

Similarly, each measurement point in a given dataset is assigned a unique point identification number (*PT_ID*). In some cases, multiple measurements are co-located, in which case, each measurement is identified by a measurement identification number (*MEAS_ID*).

The first subset of data in a selected dataset that we excluded involves measurement points and ruptures flagged as from aftershocks in FDHI database. Three events have aftershock surface ruptures and measurements: Kumamoto, Landers, and Yushu. Aftershock related features were eliminated because it is likely that aftershocks behave differently from mainshocks and produce surface ruptures that have different characteristics. Ground motion model development usually also excludes aftershocks or includes an aftershock term to capture the mainshock-aftershock difference.

For Denali, the style of faulting on the Susitna Glacier fault is reverse instead of strike-slip. Ruptures associated with this entire segment (the western most segment) were excluded. Displacement measurements on this segment are from a different source than measurements on the main surface rupture, so our rule of working with one selected source automatically eliminated them. Other strike-slip events in the database do not appear to have segments with distinctively different faulting styles.

Some mapped surface ruptures (and associated displacement measurements) were excluded because they appear to be from nontectonic processes or are triggered along other existing faults. All excluded ruptures are classified as distributed in the database. They have little to no effect on analysis of principal ruptures and displacement. Affected events include ChalfantValley (scatter ruptures in a large area to the northwest, likely non-tectonic), Kobe (ruptures and displacement measurements about 4.5 km to the southwest, appear to be triggered), and Yutian (dense features to the northwest, appear to be due to other processes).

For a given event, if the selected measurement dataset contains measurements that belong to one of the following three categories, they were eliminated from further analyses:

1. Measurements that do not have recommended net displacement values (The *net slip* field in the database is populated by “-999”).
2. Measurements that have poor quality. We excluded all measurements with quality code greater than or equal to 3000 (see Table B.3 for explanations of quality codes). Reasons for assigning a specific measurement as having quality code of 3000 or greater constitute sound reasons for excluding them. We kept all measurements with quality codes of 2000 and 2001 because these are high quality data, flagged as “check” only because they co-locate with measurements in competing datasets. This is not a problem because we worked with only one selected dataset.
3. Measurements that are co-located with other measurements in the selected dataset. These co-located measurements usually reflect measurements made through time, (i.e., with varying days elapsed after the event when the measurements were made). These events include ElmoreRanch, Napa, and SuperstitionHills. We selected the newest measurement and excluded others because (1) for most events, no information is available on when the measurements were made; (2) for other events, most measurements were made within a couple of months after the event; and (3) the only reason we would want to choose a measurement that is not the newest in the time series would be if we want to exclude displacement that is not co-seismic (e.g., after slip), which is not practical given data limitation. In the end, it does not matter because only one co-located point actually has a recommended net displacement value with a few exceptions. One exception is a pair of Napa points (*PT_ID* of 289 and 290 are co-located), and we decided to randomly select 290 and exclude 289. Another exception is a pair of Darfield points (*PT_ID* 30, *MEAS_ID* 1 and 4). It is not clear why these co-located measurements have different displacement values. *MEAS_ID* 4 was chosen because it is noted as having higher quality in the original publication.

2.2.4 Modification to FDHI Principal and Distributed Classification

Upon examining the classification in the October 2020 FDHI database, we noticed a few cases in which principal and distributed rupture and measurement classification appears inconsistent with definitions given or with our expectation, so we made slight modifications. In other cases, the FDHI classification was modified to be consistent with our modeling approach. Details of modifications for each event were discussed internally and agreed upon by team members. Also,

details of modifications were made available to the FDHI database team who, in many cases, agreed with our modifications and updated their classification in the subsequent database updates.

For most events, the FDHI classification is honored without modification. We appreciate the “ranking” attribute in the FDHI database. It provides convenience to the model development teams. More importantly, it helps achieve some level of consistency among teams participating in the FDHI project.

Aside from modifications to miscellaneous features that appear to have been assigned a questionable category, the need for modification arises from the way we calculate displacement hazard, which is based on the different behavioral characteristics of principal and distributed ruptures. By definition, principal displacements are associated with longer and relatively continuous rupture traces that show long-term geologic evidence of faulting. They have larger amplitudes. Distributed displacements are associated with scattered minor features. Their appearances can be random, although in general, they attenuate with increasing distance to principal rupture trace. In the current project, we follow the P11 approach of modeling principal and distributed displacement separately. Principal displacement is modeled as a function of along-main-trace distance, whereas distributed displacement is modeled as a function of strike-normal (off-main-trace) distance. Modification to the FDHI classified ruptures and displacements fall into three categories: (1) distributed to principal, (2) distributed to nontectonic, and (3) miscellaneous. No measurements and ruptures classified as principal in the FDHI database were modified.

We modified the FDHI classification of some distributed measurements to principal when these measurements (1) have substantial net slip (e.g., greater than a couple of meters or their net slip is substantial compared to nearby principal net slip), (2) are associated with ruptures that are relatively long and continuous (e.g., longer than a few kilometers); and in some cases, (3) are associated with faults mapped before the earthquake. When these criteria are met, observed net displacements often vary along strike, rather than in the strike-normal direction. Therefore, modeling them as principal displacement is appropriate. Modified events include ElmoreRanch, Hector, Kobe, Landers, Napa, Neftegorsk, Ridgecrest1, and Ridgecrest2.

2.3 MAIN RUPTURE TRACE AND COORDINATE SYSTEM

As discussed previously, rupture and displacement data in the FDHI database are reported at georeferenced observation locations with high position precisions. Challenges are numerous in analyzing such data in a way that takes advantage of the high location accuracy. One of these challenges, perhaps the most fundamental one, is how to define a single main rupture trace for a given event from an often-complex network of mapped surface ruptures. A single main rupture trace is needed to establish a local reference coordinate system in which measurement data are analyzed to develop predictive models. Ideally, hazards are also calculated in the same reference coordinate system relative to a mapped fault trace in application.

Figure 2.1 illustrates the reference coordinate system and geometric parameters used in P11 for a site located at distance Δ from a mapped fault trace. The footprint area to be considered has dimension z (area z^2) and is centered at the site. Distance ratio, l/L , is used to define the location along the main trace, where l is along-strike (along-main-trace) distance, and L is total rupture length. An earthquake may or may not rupture the entire fault. Parameter s represents rupture location on the mapped fault. In P11 study, principal and distributed displacements are modeled as a function of l/L and Δ , respectively.

In fault displacement studies, the amplitude of principal displacement (and total displacement or deformation across the entire rupture zone) is plotted against along-strike distance to establish an along-strike displacement distribution profile (also known as displacement curve or slip profile). Specific displacement parameters vary from study to study. For example, the displacement parameter is net slip for most FDHI teams. It is lateral slip in P11, and vertical slip in Youngs et al. (2003), Moss and Ross (2011), and Moss et al. (2022). See Table 3.1 in Kuehn et al. (2022) for more information on displacement metrics and parameters used in past and ongoing fault displacement hazard studies. Displacement profiles are useful in understanding rupture mechanics and in studying the characteristics of earthquake energy release. They are essential in fault displacement hazard assessment.

There is not a standard approach to define a general along-strike direction to plot displacement profiles. For example, different approaches were used in two influential publications. Wesnousky (2008) developed slip profiles by connecting measurement points and plotting displacement against the cumulative rupture length calculated using coordinates of measurement points. DuRoss et al. (2020) projected displacement measurements to mapped rupture traces to develop displacement curves for individual rupture traces, then combined them along a linear trace to obtain a cumulative displacement profile. Lack of a standard approach makes it difficult to compare displacement curves developed by different researchers and from different earthquakes, particularly when detailed characteristics such as displacement gradients in areas of complex rupture geometry are of interest. It also leads to inconsistent estimates of total surface rupture length which is an important parameter in hazard assessments.

Distributed displacement and deformation generally decrease as distance to the main rupture trace increases. How displacement or deformation attenuates with increasing rupture distance is another important aspect of fault displacement hazard assessment. Evaluation of the attenuation characteristics of distributed discrete displacement and off-fault deformation in the near-fault region is particularly sensitive to rupture distance, which can be affected substantially by how a main rupture trace, or a reference line is defined.

In the published PFDHA, including the P11 study, FDMs predict principal displacement as a function of along-strike distance (and other variables). They predict distributed displacement as a function of closest distance to main rupture trace. This implies a local reference coordinate system with the two orthogonal reference axes in the fault-parallel and fault-perpendicular directions, respectively. Again, there is not a standard way to define a main fault trace, leading to

inconsistency in hazard calculations at different sites along the same fault due to the complex nature of fault geometry. This problem is especially pronounced when attempting to develop fault displacement hazard maps such as those shown in Figure 11 of P11 and Figure 8 in Chen and Petersen (2011).

As shown in Figure 2.1, PFDHA requires a coordinate system with the origin on the modeled (mapped) fault and the two orthogonal coordinate axes in the local strike-parallel and strike-normal directions, respectively. In P11, the origin of the reference coordinate system for each event was anchored on a main rupture trace defined from mapped surface ruptures. The main trace was drawn manually to follow the most important principal ruptures.

For this study, we developed a method to determine the main rupture trace from mapped surface ruptures called the least-cost path (LCP) analysis. The LCP is essentially a semi-automated version of the manually drawn main trace in P11. An introduction to the LCP analysis method in the ArcGIS is included in electronic supplements. A brief summary is provided below.

2.3.1 Least-Cost Path Analysis

The LCP analysis is a raster analysis using source and destination points to find the most cost-effective path over a given surface. The “cost” is not necessarily economic based but can be a function of time, distance, or a unitless criteria defined by the user (Briney, 2014). Requirements for the LCP analysis include a source raster, cost raster, destination raster, and an algorithm to calculate the LCP (Chang, 2012).

The source raster defines the starting point of calculating the minimal accumulated travel costs to each cell within a given surface. The cost raster combines the various “costs” or impedance into one comprehensive raster. Using the source and cost rasters, a cost distance analysis is performed using the center, or node, of each cell to calculate the cost of travel to each of its eight neighboring cells. The result of this analysis is a cost distance raster representing the accumulated cost of travel to each cell from the source. The cost distance analysis also creates a backlink raster that encodes the direction of travel from each cell to its lowest cost neighbor. The cost path analysis then uses the destination raster and the cost backlink raster to trace a line back to the source following the direction of each cell to the lowest cost neighbor thus producing the least-cost path.

2.3.2 LCP as Main Rupture Trace and GC2 as Coordinate System

Using the LCP-defined trace as the main rupture trace is a substantial improvement over the manually drawn main rupture trace in P11. Manually defining a main rupture trace requires considerable professional judgement and can be hard to reproduce when mapped surface ruptures consist of numerous parallel-subparallel traces, when they form a complex crisscrossed network, when they are highly fragmented, and when large gaps exist between mapped traces either because these areas are inaccessible, or because surface ruptures are obscured by shallow soft sediments.

As an example, the LCP for Landers is shown in Figure 2.2a, along with mapped principal and distributed ruptures. Measurement points are shown in the close-up view of a stepover area (Figure 2.2b). The LCP follows the mapped rupture traces closely, particularly the principal rupture traces, and represents the best pathway connected by mapped ruptures from one endpoint of the rupture zone to the other. It retains the geometric properties (turns and bends) of the mapped rupture traces that it follows. Having the main trace closely follow mapped rupture traces is highly desirable in modeling behavior of distributed ruptures as they mostly vary with closest distance to the main rupture (i.e., strike-normal distance). However, geometric complexity in the LCP can lead to strike-normal and strike-parallel coordinates that vary sporadically. To improve distance calculations, we adopt the GC2 system developed by Spudich and Chiou (2015) to calculate the fault-parallel (u_LCP) and fault-normal (t_LCP) coordinates for all observation measurement points relative to the LCP. The u_LCP and t_LCP are then used in subsequent data analyses and model development. Coordinate u_LCP is taken as the variable l and t_LCP as the variable r in PFDHA framework of P11. Rupture length, L , is the LCP length.

2.3.3 Comparison of LCP with ECS

As mentioned previously, the FDHI database includes strike-parallel (u) and strike-normal (t) coordinates of a point with respect to ECS, for which the strike-parallel axis follows a smooth reference trace. The ECS reference trace for Landers is also depicted in Figure 2.2. The main differences in reference traces between LCP and ECS are (1) LCP is longer than ECS because ECS is highly smoothed; (2) LCP follows mapped surface rupture, and ECS trace does not; and (3) in areas very close to these reference traces, LCP divides features on its two sides in a way that is perhaps more meaningful geologically than ECS. The differences between LCP and ECS are illustrated more clearly on the zoomed-in view (Figure 2.2b).

Figure 2.3a compares along-strike distribution of measured principal displacement in the LCP and ECS coordinate systems, using Landers as an example. Along-strike distance becomes increasingly and systematically larger along LCP than along ECS because LCP is longer. This affects the along-strike gradients of estimated displacement curves somewhat, but not substantially.

The choice of coordinate system affects strike-normal distance more substantially and in an unsystematic way. Figure 2.3b compares strike-normal distribution of measured distributed displacement from Landers in the LCP and ECS coordinate systems. Distance to LCP is obviously different from distance to ECS. Displacement models developed using these two different coordinate systems will have different attenuation characteristics in the vicinity of these reference lines where the amplitude of displacement has the steepest attenuation. Therefore, it may be important that the coordinate system used in a PFDHA application is consistent with the coordinate system used in data analyses and model development.

2.3.4 Significance of LCP as Main Rupture Trace

Because the main rupture trace developed using LCP mostly follows mapped rupture traces, it is a good representation of the extent and geometry of the earthquake fault and clearly separates features on the two opposite sides of the fault. Highly smoothed reference traces such as ECS has the advantage of being simple, but ECS mixes features on the opposite sides of the fault in the near-source region, which can be undesirable for dipping faults in which attenuation characteristics in the hanging wall and footwall can be different. Applying GC2 coordinates relative to the LCP anchors the strike-parallel axis on the LCP, such that the resulting coordinates (t_{LCP} and u_{LCP}) are truly in the directions perpendicular and parallel, respectively, to a mapped rupture or a trace inferred from mapped ruptures. Because this coordinate system incorporates the geometric details of mapped ruptures, the strike-normal coordinate (t_{LCP}) is a more accurate measure of distance to the rupture trace.

The ability to capture geometric details of mapped ruptures and to use an accurate rupture distance is important in the P11 PFDHA approach. A unique feature of this approach is that it includes a component model quantifying the uncertainty in the location of surface rupture from future earthquakes. Location uncertainty is quantified by systematically measuring distances between a fault trace that geologists were able to map before an earthquake and the precise location of the corresponding surface rupture trace mapped after the earthquake. Such distance measures would not be meaningful if either the rupture traces mapped after an earthquake or fault traces mapped before that earthquake are highly simplified and smoothed.

In a PFDHA application, a hazard analyst would usually start with mapped fault traces. For faults with simple geometry and a well-defined main fault trace, distance calculations are straight forward. In such cases, LCP analysis and GC2 calculations may not be necessary. However, earthquake faults are rarely simple. In fact, they are often complex, particularly when mapped in a detailed manner with rapidly advancing remote sensing techniques. Also, mapped surface ruptures from an earthquake are often used to update and improve fault traces. Updated fault maps are then used in future hazard calculations. For example, the current USGS Quaternary Fault and Fold database (<https://doi.org/10.5066/P9BCVRCK>) incorporates surface ruptures from the 1992 Landers earthquake, which as depicted in Figure 2.2, forming a complex fault zone. Similar to applications in analyzing surface ruptures for model development, LCP can be applied to define a main fault trace from complex mapped faults for hazard analysis. In this case, the ability of LCP to follow actual mapped fault traces and to reflect turns and bends of mapped faults is especially desirable in producing large scale and detailed probabilistic fault displacement hazard maps. Producing probabilistic fault displacement hazard maps for major fault zones in California bears practical and engineering significance in many applications that are important to the missions of the CGS, California Department of Transportation, and USGS.

2.4 AGGREGATED NET PRINCIPAL DISPLACEMENT

Often earthquakes rupture the surface along multiple subparallel fault traces and branches as is the case of Landers (Figure 2.2). In areas where principal ruptures overlap along the main trace (i.e., the LCP), aggregating principal displacements on these parallel traces is necessary to obtain the total principal displacement across the rupture zone. The total principal displacement is then treated as if it is located on the main trace, and an along-strike displacement profile for the aggregated principal displacement is established for model development.

The calculation of total principal displacement would be straightforward if displacement measurement points on multiple subparallel rupture traces have common along-strike coordinates (i.e., common u_{LCP} values) and if they are spaced regularly along the LCP. This is not the case for displacements measured by geologists in the field. Figure 2.2b clearly shows that displacement measurement points are distributed irregularly along rupture traces. These points do not have common u_{LCP} values, nor do they have regular spacing.

A method that makes intuitive sense would be to resample displacement measurements by interpolation to obtain values on a set of points equally spaced along the main rupture. Such resampling, however, is undesirable for model development because (1) resampling would alter (increase) the original sample size substantially; (2) resampling would result in regression statistics dominated by interpolated values instead of the original measured values; and (3) interpolated displacement values depend on interpolation method, and it is not possible to predict which interpolation method works best prior to advanced statistical analyses for principal displacement model development.

We developed a method in which interpolation and summation are performed only at the original measurement locations so that the original sample size is maintained, and interpolation is minimized. Figure 2.4a is the map view of a simple case in which a principal rupture branch is subparallel to the main rupture trace defined by the LCP. As can be seen from this figure, both the main rupture and the branch consist of multiple mapped rupture lines, and each line has a unique identification number (RUP_ID) in the FDHI database. In our analysis, though, both the main rupture and the branch are treated as continuous surface ruptures (either a segment or a branch). Small scale discontinuities and complexities seen in Figure 2.4a are common and are often an inherent part of surface rupture due to near surface and, sometimes, surface constraints. A mapped rupture trace may end artificially as the result of an area being inaccessible to the geologists. There are cases where gaps between two rupture lines are so small that the lines are practically continuous. Yet, they are mapped as separate lines and have separate RUP_IDs . For these reasons, we do not use RUP_ID in our analysis. Instead, we manually group mapped ruptures into segments or branches, assign a number to each, and use these numbers to track overlaps along the LCP.

The method to aggregate principal displacement involves manually identifying subparallel segments and branches that overlap when projected onto the LCP to develop a simple segmentation

model, attributing each overlapping segment or branch with a unique identification number, then utilizing the identification number to aggregate principal displacement with an algorithm.

2.4.1 Simple Segmentation Models

The purpose of developing a simple segmentation model is to identify and organize measurement points associated with subparallel segments and branches. Segmentation is based on information in published geologic literature; finite-fault models in the USGS Significant Earthquake Archive (USGS, 2021) and in the SRCMOD online database of finite-fault rupture models (Martin and Thingbaijam, 2014); and characteristics of mapped surface ruptures. Stepovers, substantial changes in orientation, large gaps on the rupture-length scale, presence of low valleys on the along-strike slip profile, and the occurrences of distributed ruptures at the rupture length scale are good indications of segment boundaries. For each event, we assigned sequential whole numbers to segments. The segment number increases from the west to the east (i.e., in increasing u_{LCP} direction). Numbers with decimal points indicate subparallel branches, and the single digit after the decimal point indicates branch number. As an example, Figure 2.5 illustrates the segmentation model for Landers. Each measurement point is assigned a segment or a branch number. These numbers are then used to track principal displacements that should be added to obtain aggregated principal displacement.

2.4.2 Computation of Aggregated Principal Displacement

The method for aggregating principal displacement is explained using the simple case of a rupture branch subparallel to the main rupture shown in Figure 2.4a. The figure is a map view of principal ruptures, distributed ruptures, displacement measurement locations, and the main rupture trace delineated by the LCP. Locations of measurement points are shown as solid blue dots, darker blue on the main trace and lighter blue on the branch. The main trace shown is a small section of segment 1, and the branch is labeled as 1.1 in the simple segmentation model for Landers showing in Figure 2.5.

As discussed previously, principal displacement is modeled as a function of distance along the LCP, which essentially means relocating (or projecting) all measurement points to the main trace according to their u_{LCP} coordinates, as indicated by the thin gray lines (perpendicular to the main trace) with arrow heads and red open circles in Figure 2.4a. Principal displacements are plotted against the u_{LCP} coordinates in Figure 2.4b. Connecting measurement points on the main trace forms displacement profile for segment 1. Similarly, connecting measurement points on the branch forms the displacement profile for branch 1.1.

On displacement profile for segment 1, an interpolated value is obtained at each u_{LCP} coordinate where there is a measurement on branch 1.1. These interpolated values are calculated by linear interpolation using the two nearest neighboring points on segment 1. Likewise, on the

displacement profile for branch 1.1, an interpolated value is obtained at each u_{LCP} coordinate where there is a measurement on segment 1; interpolated values are calculated by linear interpolation using the two nearest neighboring points on branch 1.1.

Finally, at each u_{LCP} value where there is a measurement point, the aggregated principal displacement is the arithmetic sum of the measured and interpolated displacement values. The total principal displacements are shown as open red circles in Figure 2.4b. These are the values used in subsequent regression analysis for principal FDM development.

Figure 2.6a is a map view covering the complete length of branch 1.1 that overlaps with segment 1 along the LCP, and Figure 2.6b shows individual principal displacement measurement points and aggregated displacement. Measurement points in Figure 2.6a are color coded by their percent contribution to the aggregated principal displacement. The percent contribution data will be useful in providing guidance for hazard calculation in engineering application.

2.4.3 Comparison of Aggregated and Measured Displacement Profiles

Figure 2.7 compares the along-strike distributions of the as-measured and aggregated principal displacements for Landers as an example. Measured displacement profile shows higher along-strike variability with these general patterns: (1) displacement is higher in the central portion of rupture and decreases toward ends of ruptures, and (2) there are areas of apparently low displacement, which seem to correspond well with segment boundaries. Distribution of the aggregated principal displacement is less variable compared to the case when displacements are not aggregated and simply projected onto the LCP trace. The aggregation process reduces variability and is a better representation of total slip on all principal ruptures.

As mentioned in Chapter 4, we adopt the PFDHA framework of P11 and model principal displacement and distributed displacement separately. Therefore, the displacement aggregation across the rupture zone is only applied to the principal displacement. Other model development teams participating in the FDHI project aggregate not only principal displacement, but also distributed displacement for their model development. The aggregation method varies from team to team. Perhaps the similarity can be expected considering how small distributed displacements are compared to principal displacements and the small differences between along-strike displacement distribution in the LCP and in the ECS coordinates as shown in Figure 2.3a.

2.5 DATA DISTRIBUTION

Figure 2.8 shows the geographic distribution of epicenters of all strike-slip earthquakes in the FDHI database, in which events included and excluded from our model development are identified. Figure 2.9 shows the distribution of selected displacement measurements (D) plotted against \mathbf{M} in along-strike distance bins. Each bin is defined by normalized location (l/L) along the main trace. It shows displacement generally increases with \mathbf{M} , and there appears to be larger scatter for smaller

M. Tier-1 and tier-2 data are explained in Section 3.2.2.1 and in Appendix D. Figure 2.10 shows distribution of displacement measurements along normalized rupture location for individual selected earthquakes. There is a large variation in spatial coverage of displacement measurements and in sampling rates. Figure 2.11 is a histogram showing number of slip measurements (N_{slip}) for each individual earthquake. Out of the 29 strike-slip earthquakes, four have more than 200 measurements, and all four earthquakes are of $M \geq 7.1$. These four earthquakes are the 1992 Landers ($N_{\text{slip}} = 566$), 1999 Izmit Kocaeli ($N_{\text{slip}} = 263$), 2013 Balochistan earthquake ($N_{\text{slip}} = 247$), and 2019 Ridgecrest mainshock ($N_{\text{slip}} = 226$).

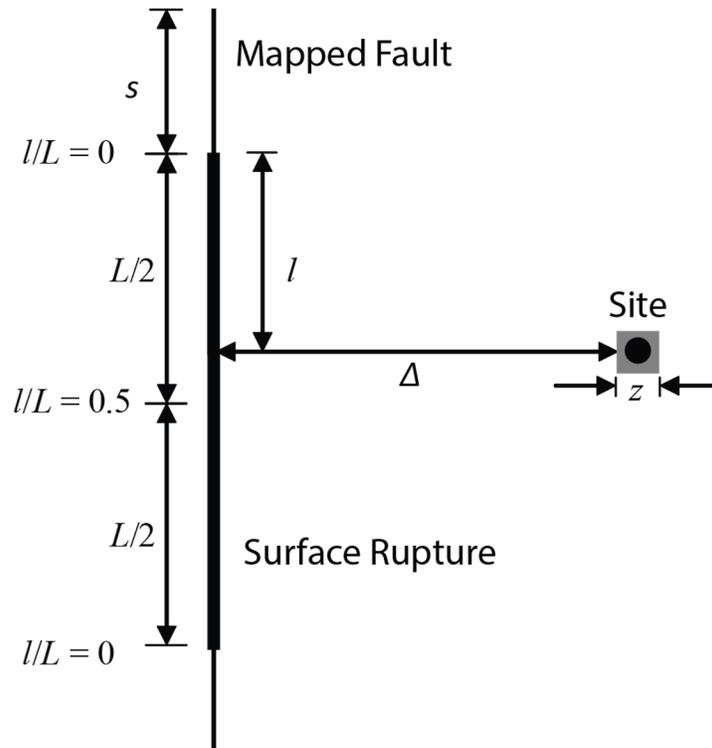
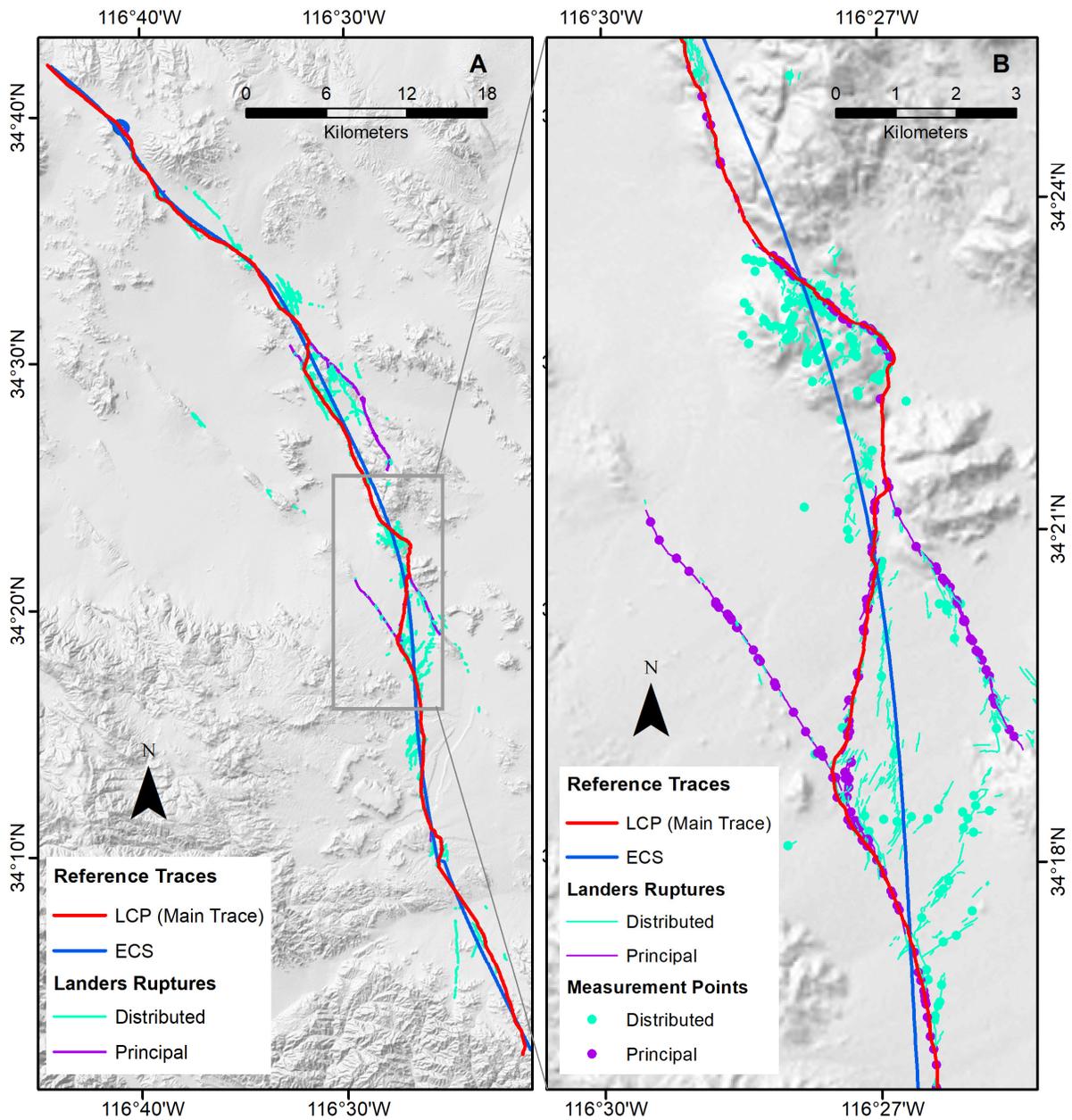


Figure 2.1. Geometric parameters used in fault displacement model development and hazard analysis (modified from Petersen et al., 2011). Variables l and Δ are utilized to specify the along-main-trace and off-main-trace position, respectively, of a point. Variable L is the length of the main trace. Variable s is utilized to track the position of main trace along the mapped fault trace.



Basemap: ESRI

Figure 2.2. Comparison of least-cost path (LCP) and event coordinate system (ECS) reference lines for the 1992 M 7.28 Landers earthquake. (A) View of the entire rupture length. (B) Close-up view of a stepover area.

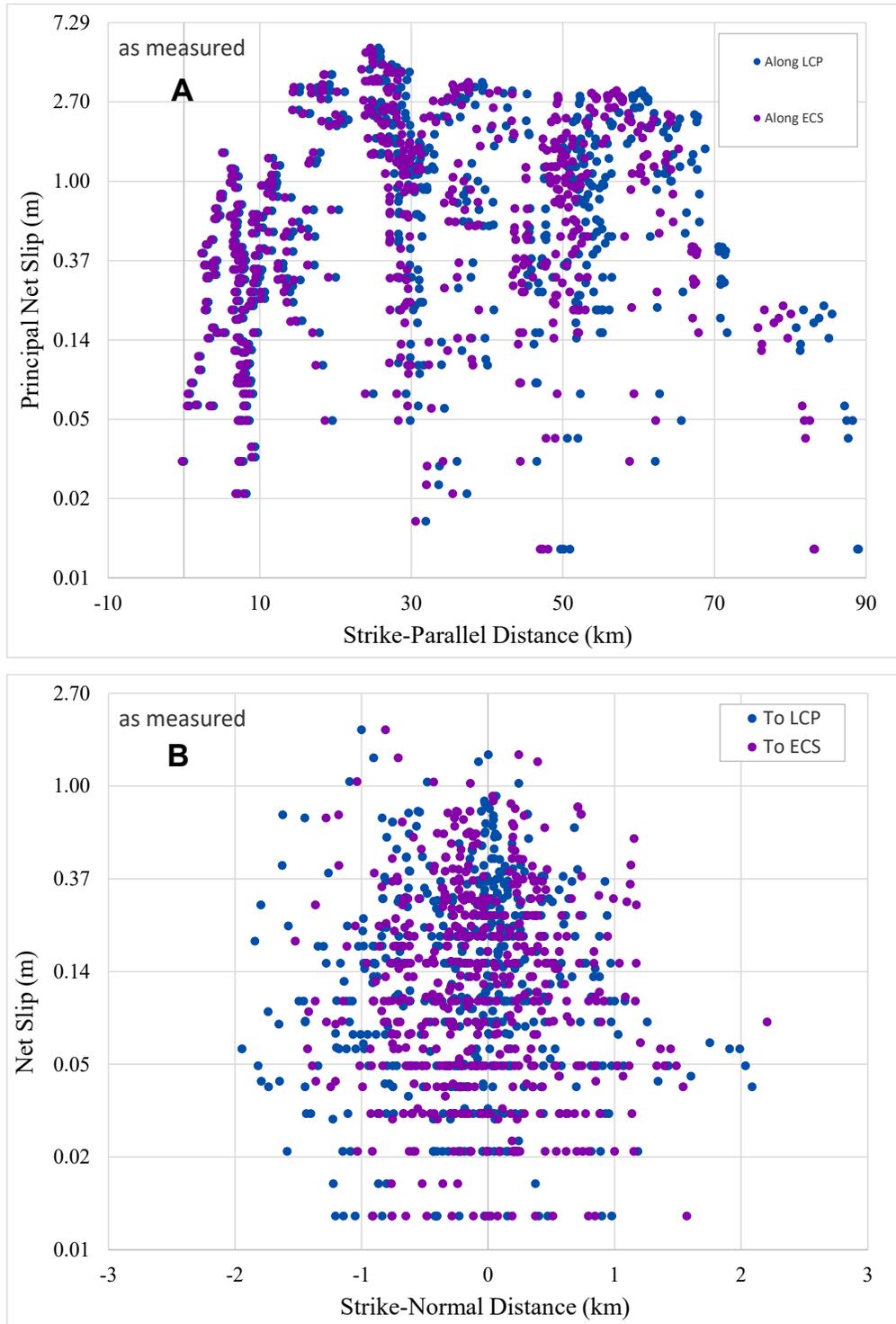


Figure 2.3. Comparison of measured displacement from Landers in the coordinate systems for least-cost path (LCP) and event coordinate system (ECS). (A) Along-strike distribution of principal displacement. (B) Strike-normal distribution of distributed displacement.

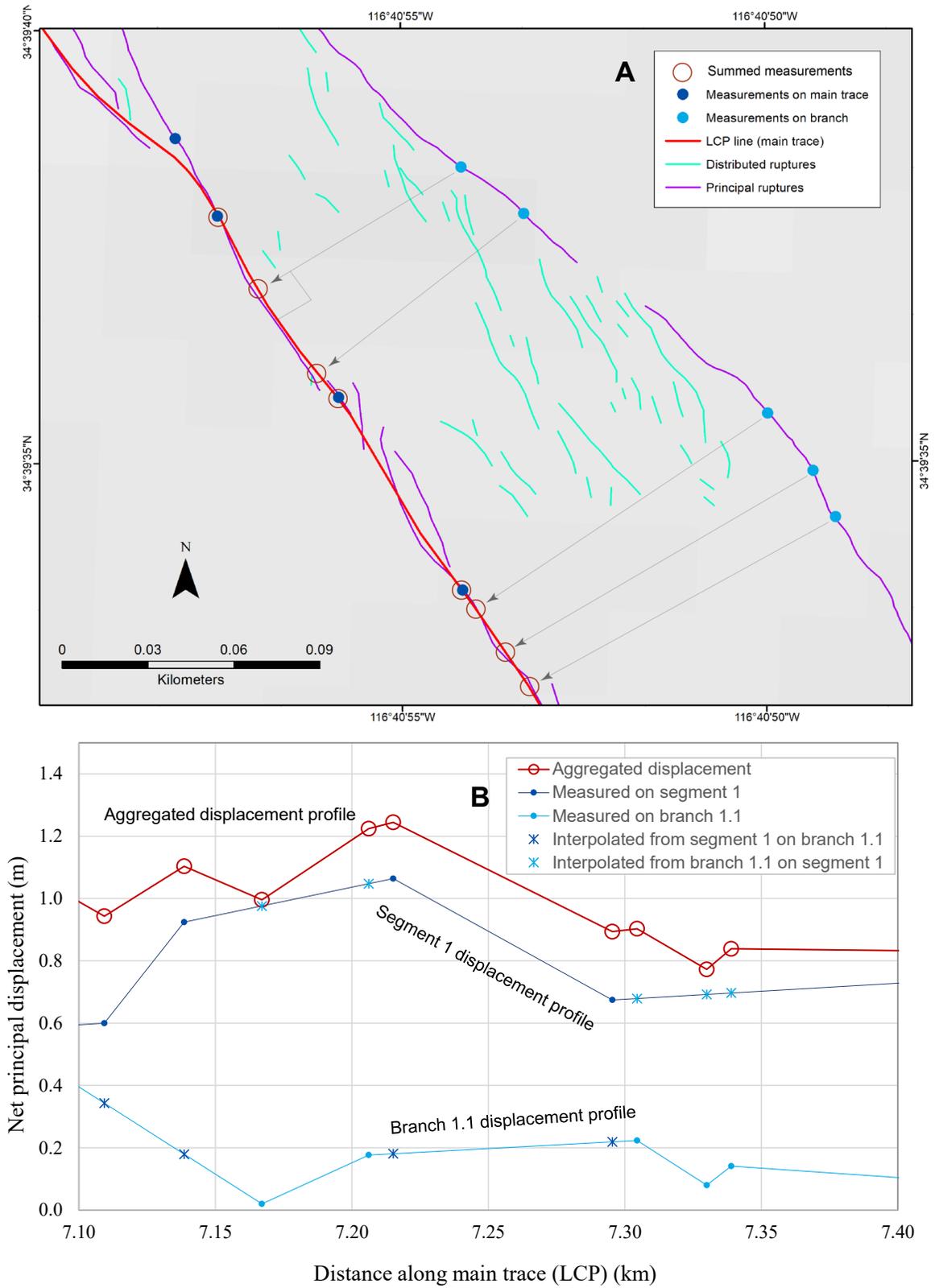


Figure 2.4. Aggregation of principal displacements on subparallel principal rupture traces. (A) Map view. (B) Individual and aggregated displacement.

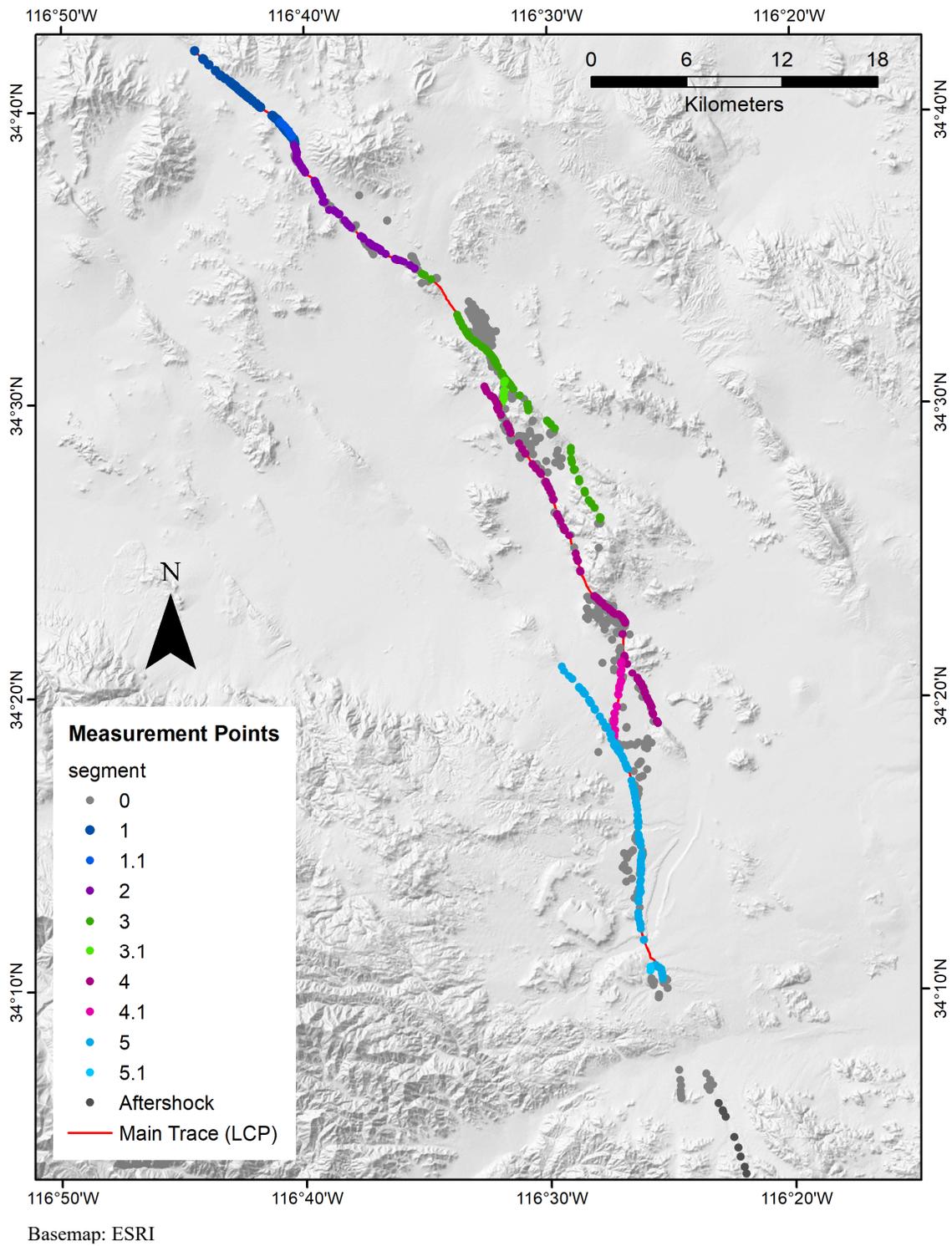


Figure 2.5. Segment and branch number assignment for Landers measurement points.

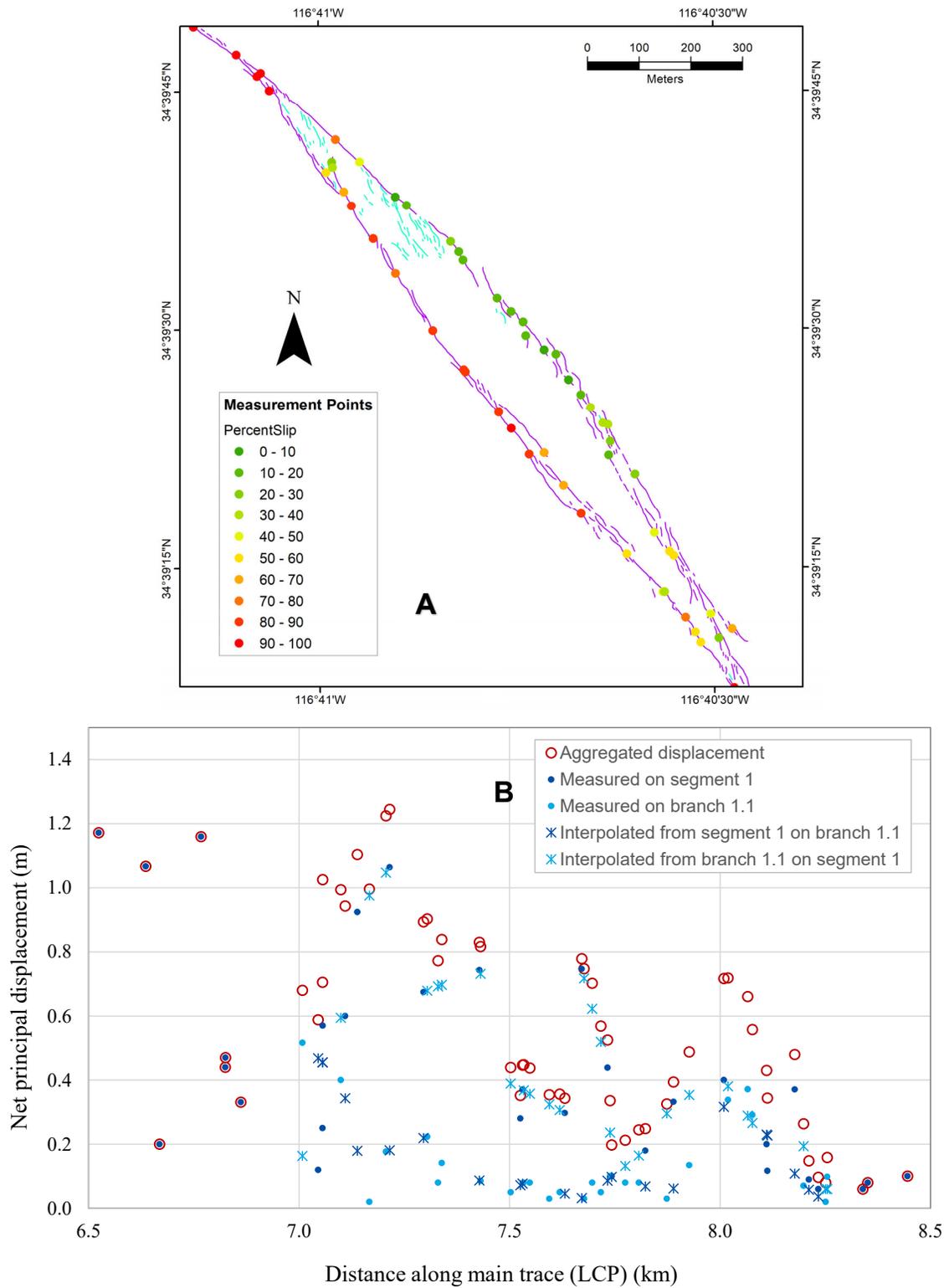


Figure 2.6. Aggregation of principal displacement for segment 1 and branch 1.1 of Landers surface rupture. (A) Map view with measurement location color-coded by percent contribution to the aggregated displacement. (B) Individual and aggregated displacement.

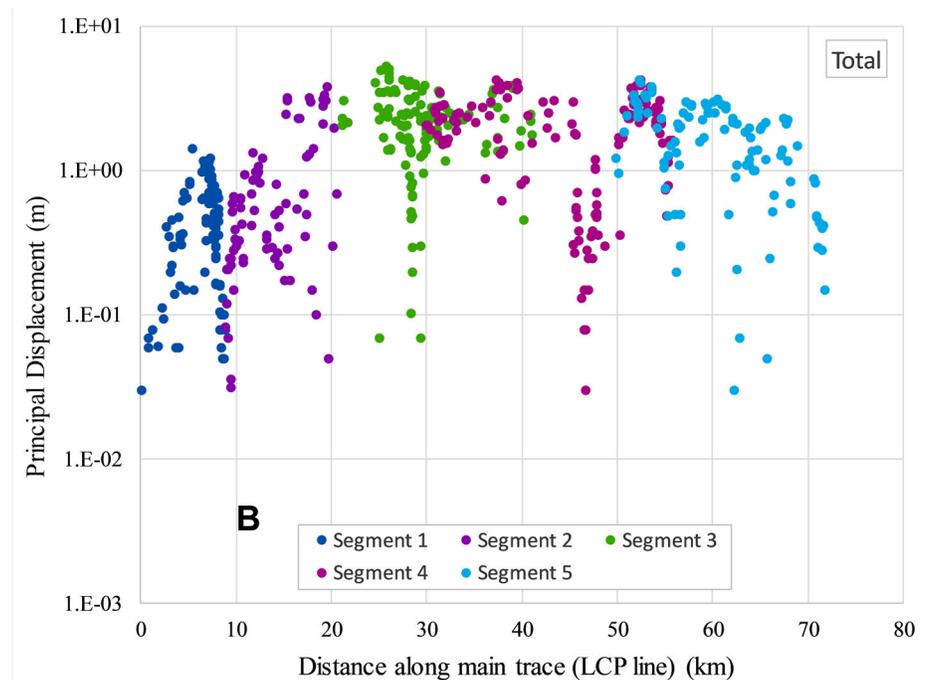
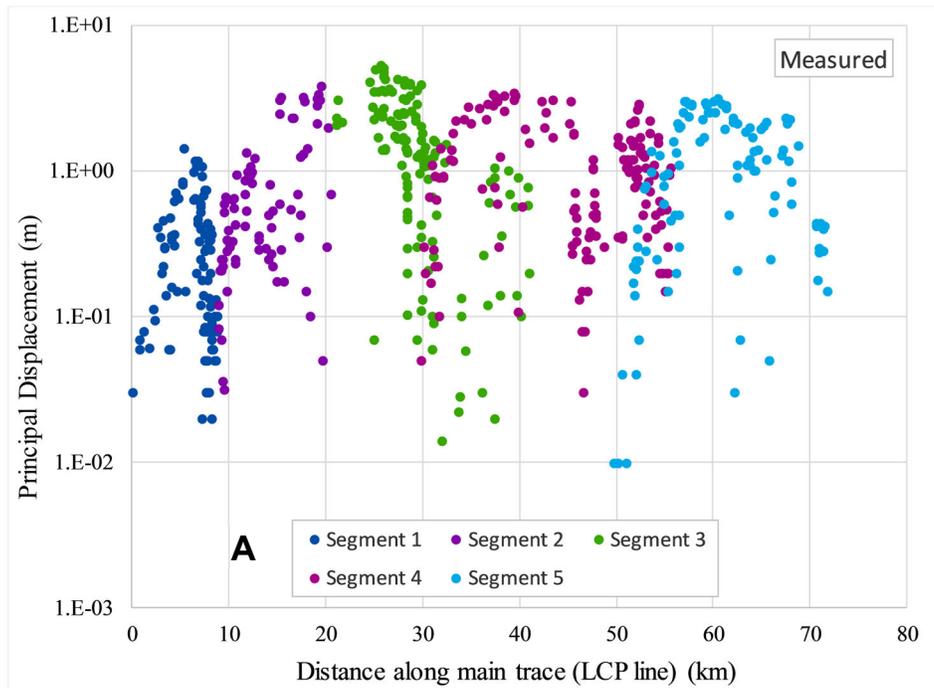


Figure 2.7. Comparison of along-strike distributions of principal displacement. (A) As-measurement principal displacement. (B) Aggregated principal displacement.

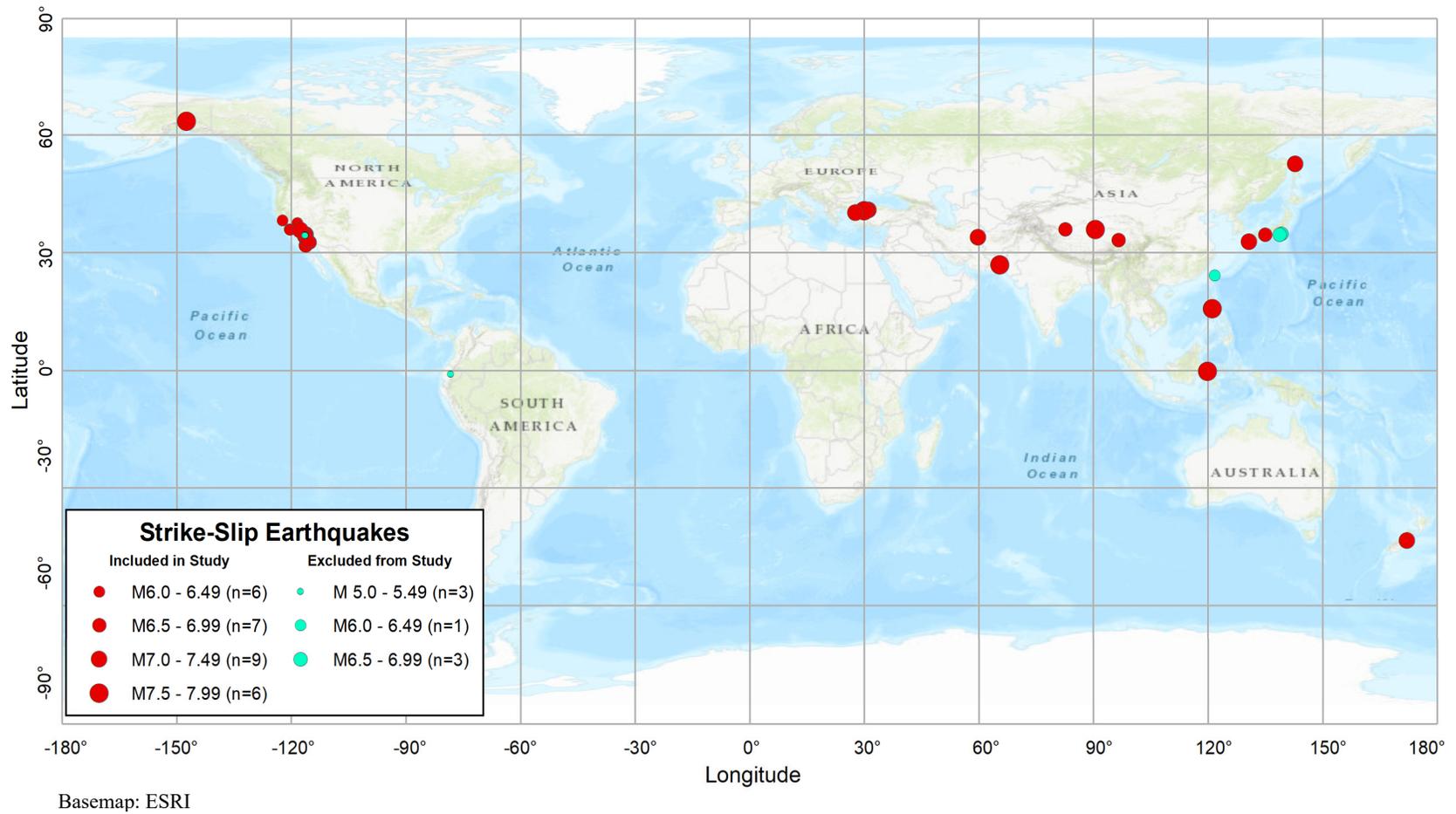


Figure 2.8. Epicenter distribution of strike-slip earthquakes in the Fault Displacement Hazard Initiative (FDHI) database (Sarmiento et al., 2021), including earthquakes selected for model development and those excluded for reasons discussed in Section 2.2.1.

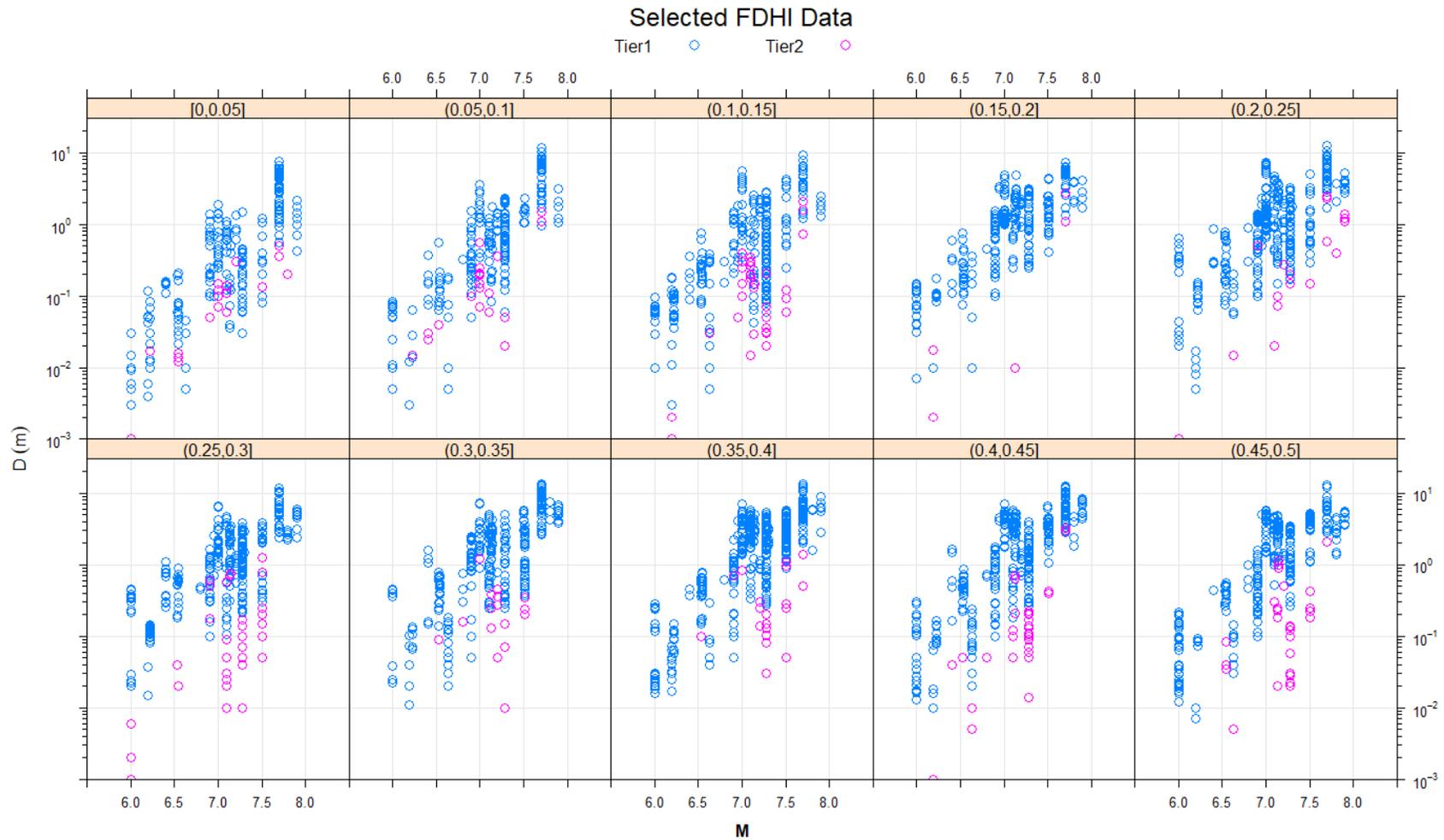


Figure 2.9. Distribution of displacement measurements plotted with magnitude (M) in along-strike distance bins. Each bin is defined by normalized location along the main trace (i.e., values of l/L , where L is rupture length and l is distance to the nearest end of rupture). Tier-1 and tier-2 data are explained in Section 3.2.2.1 and in Appendix D.

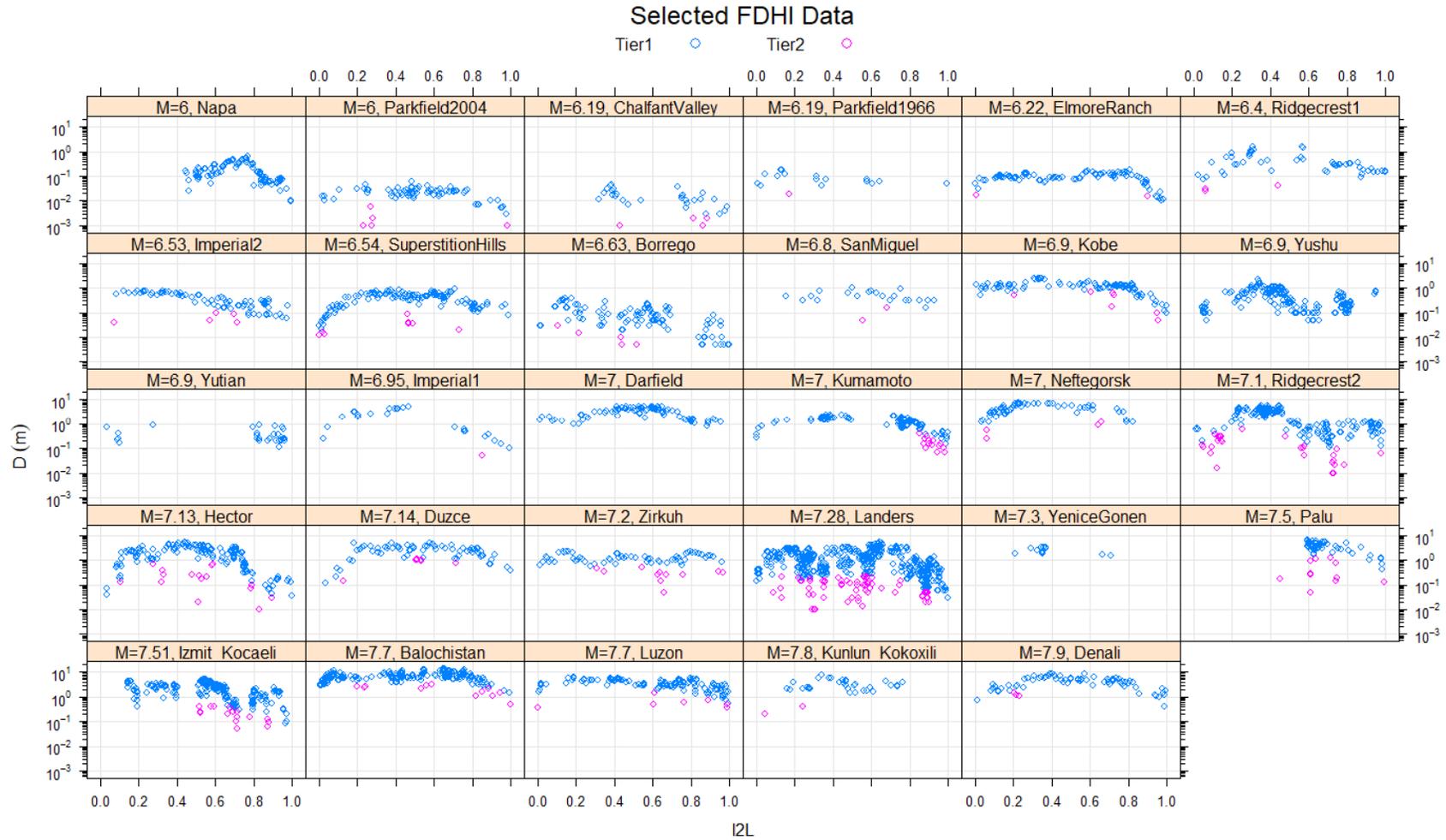


Figure 2.10. Distribution of displacement measurements along normalized rupture location (l/L) for individual earthquakes. Tier-1 and tier-2 data are explained in Section 3.2.2.1 and in Appendix D.

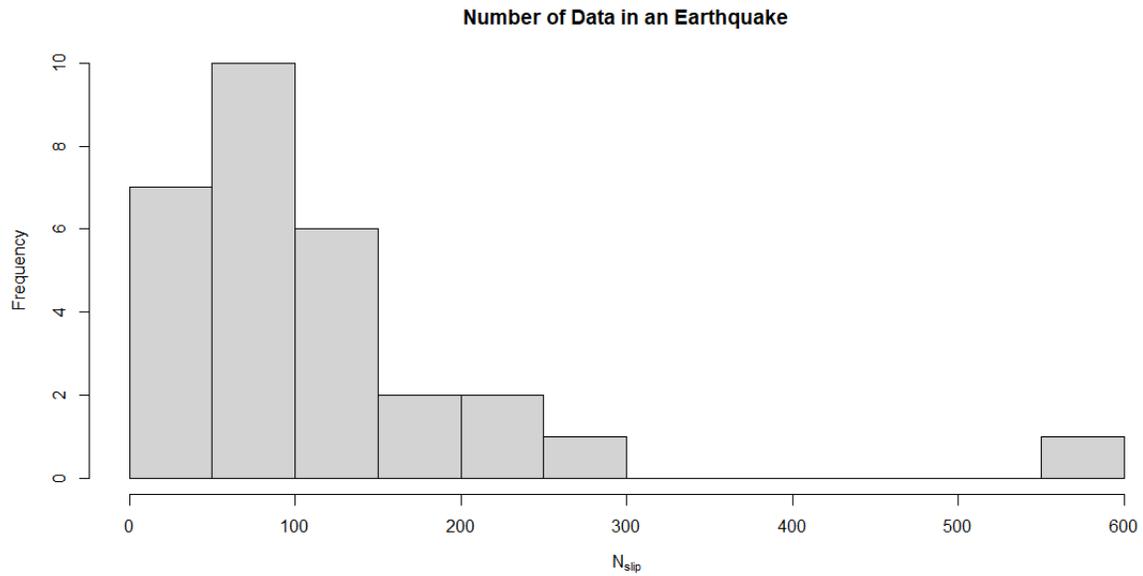


Figure 2.11. Histogram of the number of slip measurements (N_{slip}) for an individual earthquake. Out of the 29 strike-slip earthquakes, four have more than 200 measurements, and all four are of **M** 7.1 or larger. These are the 1992 Landers ($N_{slip} = 566$), 1999 Izmit Kocaeli ($N_{slip} = 263$), 2013 Balochistan ($N_{slip} = 247$), and 2019 Ridgecrest mainshock ($N_{slip} = 226$).

3 Development of Probabilistic Distribution Models for Principal Displacement

As stated in the Introduction Chapter, our study aims to develop an updated probabilistic distribution model for the aggregated principal net displacement at points along the main surface trace (FDM, for brevity) of a strike-slip earthquake. FDM is estimated using a distributional regression on the selected FDHI displacement data described in Chapter 2. We adopt simple improvements to P11's model formulation so that hazard analyst can readily use our model to compute the probability of exceeding a specific value of principal displacement, given the magnitude (\mathbf{M}) of a future surface rupture and the normalized site location ($l2L$) along the main trace of that rupture. This probability of exceedance is one of the required input models in PFDHA, as described in Chapter 4 and Appendix A.

Our model development begins with a simple update to the FDM of P11 (their elliptical model, to be exact). This update provides us an opportunity to identify crucial refinements to P11's original model formulation. Once the identified refinements are incorporated, modeling of the non-normal probability distribution of FDM's response variable becomes the main focus of subsequent model development, which leads to three interim updates to P11 before settling on our final model. In the first interim update, we use a data trimming method to help reduce the severity of non-normality in response variable. We then venture outside the familiar world of normality and try out the skew-normal (Azzalini, 1985; Azzalini and Capitanio, 2014) and the skew- t (Azzalini and Capitanio, 2003; Azzalini and Capitanio, 2014) distributions, which afford larger flexibility than the normal distribution by regulating the shape of their density functions through one or two additional distribution parameters. The simple update and the three interim updates are summarized in Section 3.2. The last model update, which assumes that response variable follows a negative exponentially modified Gaussian (nEMG) distribution, is presented in Section 3.3. In Section 3.4, we compare all FDMs presented in this Chapter (additional comparisons in the hazard domain are provided in Chapter 4) and explain why we prefer the last update. In Section 3.5, we assess the epistemic uncertainty in the predicted quantiles resulting from the epistemic uncertainty in the \mathbf{M} -scaling relation of our preferred model. Finally, additional discussions about the preferred model are presented in Section 3.6, including a justification of the use of (symmetric) ellipse equation for the variation along the main trace, extension to asymmetric along-trace variation, magnitude scaling relations for the average (D_{ave}) and the maximum (D_{max}) displacement over

measurement points along the main trace, similarity of the stochastic representation of selected displacement metric between our preferred model and other published models, and the applicable magnitude range of our preferred model.

3.1 RESPONSE VARIABLE AND ITS PROBABILITY DISTRIBUTION

An important decision by the developers of an FDM is the selection of the response variable to be modeled by their FDM. This decision entails the choices of a displacement metric and its algebraic transformation.

Different types of displacement metrics have been modeled in previously published FDMs. Youngs et al. (2003) and Moss and Ross (2011) modeled the vertical displacement in their FDMs for normal earthquakes and reverse earthquakes, respectively. P11 modeled the lateral displacement in their FDMs for strike-slip earthquakes. We choose to model the net principal displacement in this study. Note that our chosen metric differs from the metric of total displacement in that principal displacement does not include contributions from secondary (distributed) displacements and off-fault continuous deformation.

Once the displacement metric D is selected, a transformation of the selected metric may be applied. In previously published FDMs, two types of transformation have been used: the identity transformation and the logarithmic transformation. The latter can be either the base-10 logarithm or the natural logarithm; the natural logarithm will be used throughout this study.

The selected transformation often dictates the type of probability distribution for use in regression analysis to model the probabilistic distribution of the response variable. For the identity transformation, the response variable is the displacement metric D itself. Because surface displacement is positive in value, the candidate probability distribution for D or its normalized form D/D_{ave} , where D_{ave} = the average of D over points along the main trace, is limited to those supported on the positive half of the real line $(0, \infty)$. As two examples, gamma and Weibull distributions were used in the modeling of D/D_{ave} in Youngs et al. (2003), Moss and Ross (2011), and Takao et al. (2013).

The logarithmically transformed D can take on either positive or negative value; therefore, probability distributions with support on the entire real line $(-\infty, \infty)$ are eligible. In some of the FDMs of P11, the logarithmically transformed displacement metric was selected as the response variable, for which a normal distribution was assumed. We follow P11 and use the logarithmic transformation, but we are not limited by the normal distribution assumption, for reasons to be explained below.

In summary, our choice of the response variable is the logarithmically transformed D , where D is the aggregated principal net displacement. In forward application, logarithmic transformation is reversed via exponentiation. Hence, our FDM does not predict negative displacement, unlike when D is modeled as a normally distributed variate.

3.2 INTERIM UPDATES

In the course of our model development, we advance through four FDMs before settling on the final model. These four FDMs are interim updates to P11 that not only help improve P11's model formulation but also help shape our views on the optimal probabilistic distribution of the FDHI displacement data. Summarizing the main points of each of these interim updates is thus not only informative but also necessary. Furthermore, a comprehensive documentation of interim updates has the potential benefit of allowing hazard analysts to assess each update in hazard applications where alternative assumption on probabilistic distribution is desired.

The four interim updates are summarized in this Section. Reading through a lengthy description of interim FDMs may be tedious and unnecessary for some readers. A short summary is thus provided below so that, at first reading, readers not interested in the technical details of interim updates may read this summary and skip to the final model described in Section 3.3.

The progression of updates begins with a simple update of P11's FDM to the selected FDHI dataset (Section 3.2.1). This simple update helps us identify several necessary improvements to P11's original formulations. The identified improvements include the adoptions of bi-linear \mathbf{M} scaling relation, random-intercept mixed-effect modeling, \mathbf{M} -dependent variance of the random intercepts, and non-normal distribution of $\ln(D)$.

The first three of the identified improvements are implemented in Section 3.2.2. They bring the predicted 95th percentile to a level somewhat commensurate to historically observed maximum displacement in large earthquakes. As a way to handle data non-normality noted in Section 3.2.1.3, non-normality is made less severe through removal of a small number of data whose residuals are incompatible with the normal assumption. To facilitate identification, we devise a method using results of quantile regression (Koenker, 2005) to group data into two tiers; see Section 3.2.2 and Appendix D for more details. A total of 244 (out of 3,334) displacement data are identified as tier 2. Tier-2 data are then excluded from the regression analysis of Section 3.2.2, in favor of a closer fit to tier-1 data by the normal distribution. Despite the success in reducing non-normality, there is still non-negligible deviation of the residual distribution from assumed normal distribution (see residual diagnostic plots in Section 3.2.2.6).

To further improve fits to tier-1 data, normal distribution is replaced by the skew-normal (SN) distribution in Section 3.2.3. SN distribution is a generalization of normal distribution to introduce asymmetry in the probability density function (Azzalini, 1985; Azzalini and Capitanio, 2014). SN includes a third distribution parameter that regulates both skewness and kurtosis. The extra flexibility helps substantially improve fit to tier-1 data residuals on both the left and the right tails (see residual diagnostic plots in Section 3.2.3.5).

The SN-based model is incomplete in that it does not model tier-2 data. SN distribution struggles fitting tier-2 data when they are brought back into regression. Our aspiration to have a complete model for both tiers of data leads to the third update. In the third update, we seek an even

more flexible distribution and settle on the four-parameter skew- t distribution (ST) (Azzalini and Capitanio, 2003; Azzalini and Capitanio, 2014), which generalizes Student's t distribution by introducing non-zero skewness (and hence asymmetry in the probability density function). The ST-based model is presented in Section 3.2.4. It only slightly improves on the already good fit of SN-based update, but accommodates both tier-1 and tier-2 data (see residual diagnostic plots in Section 3.2.4.4).

One drawback of the SN and ST distributions is that they are more difficult to interpret and more cumbersome to use than the normal distribution. A second drawback is that the choice of either distribution lacks a geological underpinning. Our final and preferred model (presented in Section 3.3) improves on the above two shortcomings.

3.2.1 *Simple Update: Repeating P11 Regression Analysis on Selected FDHI Data*

Our first update to P11's FDM is named 'simple update' because it is a repeat of P11's regression analysis using the selected FDHI data. We focus on updating P11's elliptical model for $\ln(D)$ (Equation (13) in P11) because it renders a profile of D having approximately a linear slip gradient toward both edges of main trace, whereas P11's quadratic and bilinear models do not. More discussions about the shape of the predicted profile of D are given in Section 3.6.1. On the statistical side, the elliptical model has the advantage that it affords a straightforward geometric interpretation of its coefficients (see also Section 3.2.2.2).

3.2.1.1 *Models for the location parameter μ and the scale parameter σ*

In P11, response variable $\ln(D)$ was assumed to follow a normal distribution with mean (location parameter) μ and standard deviation (scale parameter) σ ,

$$\ln(D_{ij}) \sim N(\mu_{ij}, \sigma^2) \tag{3.1}$$

$$\mu_{ij}(\mathbf{M}_i, l2L_{ij}) = c_0 + m_1 \mathbf{M}_i + c_1 x_{ij}^*$$

where index j refers to a measurement in earthquake i , $x_{ij}^* = \sqrt{1 - \frac{(l2L_{ij}-0.5)^2}{0.5^2}}$, $l2L_{ij} = l_{ij} / L_i$, and L_i is the length of the main trace of earthquake i . Note that the definition of l_{ij} differs between P11 and this study. In P11 (P11's Figure 1 and our Figure 2.1), l_{ij} for measurement point j is the distance between its closest point on the main trace and its closer edge of the main trace, measured along the main trace. P11's definition of l_{ij} amounts to combining measurements symmetrically with respect to the midpoint of the main trace. Such operation is also called 'data folding.' The range of folded $l2L_{ij}$ is $[0, 0.5]$. In this study, l_{ij} is taken as the generalized strike-parallel coordinate (see Section 2.3.2) relative to the western end of the main trace. The range of

corresponding $l2L_{ij}$ is $[0, 1]$. To distinguish the two versions of $l2L_{ij}$, we denote P11's version as $l2L_{f,ij}$, where $l2L_{f,ij} = l2L_{ij}$ if $l2L_{ij} \leq 0.5$, and $l2L_{f,ij} = 1 - l2L_{ij}$ if $l2L_{ij} > 0.5$. It does not matter which version is used in the calculation of x_{ij}^* . We prefer $l2L_{ij}$ for its potential use in extending the symmetric μ to an asymmetric one.

In P11's formulation, each observation point has its own mean μ_{ij} , modeled as a linear function of covariate \mathbf{M}_i . The dependence of μ_{ij} on $l2L_{ij}$ is via the term $c_1 x_{ij}^*$, which tracks the upper half of an ellipse centered at $l2L = 0.5$. The variance parameter σ^2 is a constant for all observations (homoscedasticity), independent of \mathbf{M} and $l2L$.

This simple update is intended to be a repeat of P11's statistical analysis; hence, we follow their approach of estimating the unknown coefficients (c_0, m_1, c_1) of Equation (3.1) using least-squares regression method without earthquake effect on the intercept c_0 , instead of the preferred mixed-effect regression method to be discussed later. The estimated coefficients are tabulated in Table 3.1. The variance parameter σ^2 is estimated as the residual variance, which is tabulated in terms of standard deviation in Table 3.1. Residual of the simple update is defined as $\varepsilon_{ij} = \ln(D_{ij}) - \bar{\mu}_{ij}(\mathbf{M}_i, l2L_{ij})$, where $\bar{\mu}_{ij}$ is the estimated mean parameter for $(\mathbf{M}_i, l2L_{ij})$ computed using the estimated coefficients. Residual ε_{ij} is to be differentiated from the normalized quantile residuals presented later.

3.2.1.2 Predictive Distribution from the Simple Update

Predicted quantiles of displacement from the simple update and the original P11 FDM are shown on Figure 3.1. The 0.95 quantiles predicted for $\mathbf{M} > 8$ ruptures at location $l2L = 0.5$ are much larger than the maximum displacements observed in several large historical earthquakes. The maximum displacement of strike-slip earthquakes in the FDHI database is 13.6 m from the \mathbf{M} 7.7 Balochistan earthquake. Rodgers and Little (2006) reported a dextral slip of 18.7 m in the rupture of the 1855 New Zealand earthquake (whose magnitude is thought to be greater than 8.1) on the Wairarapa fault. The study by Kumar et al. (2006) reported 26 m of slip along the Indian Himalaya (probably associated with the 1505 \mathbf{M} 8.2 Nepal earthquake, a continental reverse earthquake). On the theoretical side, dynamic rupture modeling by Wang and Goulet (2022) yields a maximum displacement in the range of 10 to 20 m for \mathbf{M} 8 to \mathbf{M} 8.2 strike-slip ruptures.

3.2.1.3 Residual Diagnostics

A set of five residual diagnostic plots are presented on Figure 3.2. The histogram of the standardized residual (ε_{ij}/σ) is far from the standard normal distribution. Relative to the standard normal distribution, the residual histogram has a thinner tail on the positive side and a fatter tail on the negative side, similar to that implied by the residuals shown on Figure 6 of P11. This is expected because the simple update involves only a change of dataset; no changes are made to the model formulation.

3.2.1.4 Improvements to P11's Model Formulation

In this section we further analyze the residuals to help identify areas where improvements to P11's model formulations are needed. To help explain them, we invoke the mean earthquake residual $\bar{\varepsilon}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \varepsilon_{ij}$, where n_i is the number of displacement measurements in earthquake i . The quantity $\bar{\varepsilon}_i$ has been conveniently used to infer the random effect on individual earthquake's mean parameter if the adopted regression model does not include an earthquake-specific random effect on μ model's intercept. In the following discussions, $\bar{\varepsilon}_i$ is utilized to justify the bilinear magnitude scaling of the mean parameter μ (item 1 below) and to compute the within-earthquake residual ($\varepsilon_{ij} - \bar{\varepsilon}_i$). The quantity $\bar{\varepsilon}_i$ is a temporary patch to remedy the limitation of least-squares regression; it becomes obsolete once we formally adopt the mixed-effect regression in the subsequent interim updates.

1. The $\bar{\varepsilon}_i$ shown on Figure 3.3 has the shape of a wedge, indicating that μ can be modeled more closely by a bilinear function of \mathbf{M} rather than the linear function used in the simple update to P11. We fit the functional form of the ground motion model (GMM) of Chiou and Youngs (2008, 2014) to $\bar{\varepsilon}_i$. The fitted curve (red curve on Figure 3.3) tracks the magnitude trend of $\bar{\varepsilon}_i$ well, indicating that this functional form is viable for the \mathbf{M} -scaling of surface displacement. The formal adoption and implementation of the bilinear \mathbf{M} -scaling relation in the framework of random-intercept mixed-effect regression is discussed in Section 3.2.2.2.
2. Linear least-squares regression used in the estimation of the simple update (as well as the original P11) poses an estimation problem when measurement data size is uneven among earthquakes. Specifically, magnitude scaling rate may be overly influenced by a few well-measured earthquakes. With the adoption of the bilinear \mathbf{M} -scaling relation in our study, this problem is non-negligible because the four strike-slip earthquakes having at least 200 displacement measurements are all of \mathbf{M} 7.1 or larger (Figure 2.11). These four earthquakes, if not properly weighted, will have a large influence on the estimated magnitude scaling rate for $\mathbf{M} > 7.1$, a range of great interest in application. The same issue has been recognized and addressed in the ground-motion model development. To provide a statistically sound estimation, earthquake-specific random effect on μ model's intercept (Abrahamson and Youngs 1992) has been used to tame the influence of well-recorded earthquakes on estimated mean parameter. Using residuals from the simple update, an example of the impact of proper weighting of data for individual earthquake is shown on Figure 3.3. Two very different bilinear \mathbf{M} -scaling relations are revealed by the least-squares regression and the random intercept mixed-effect regression. The fit by least-squares regression indicates a linear relation, whereas the fit by mixed-effect regression indicates a bilinear relation that tracks closely the $\bar{\varepsilon}_i$ of individual earthquake. For the above reasons, we prefer mixed-effect

regression over least-squares regression for the FDM development. The mixed-effect formulation is introduced in Section 3.2.2.2.

3. The standard deviation of within-earthquake residual of tier-1 data (tier classification is discussed in Section 3.2.2.1) is not a constant along the main trace (Figure 3.4), and a heteroscedastic σ^2 is warranted. Reformulation from constant (homoscedastic) to heteroscedastic σ^2 is introduced in Section 3.2.2.3.
4. Finally, the histogram and the normal quantile-quantile (Q-Q) plot of standardized residual (Figure 3.2) indicate a probabilistic distribution that deviates substantially from the standard normal distribution, indicating the normality assumption used in the simple update is inadequate. Efforts to find probability distributions that capture the observed data non-normality are presented in Sections 3.2.2 to 2.2.4 and in Section 3.3.

Data residuals do not exhibit a trend with x_{ij}^* (not shown) and $l2L_f$ (top left panel of Figure 3.2; see also Figure 3.4). We thus conclude that the ellipse formulation (the $c_1 x_{ij}^*$ term) does not need reformulation. This conclusion is also valid for the subsequent model updates. We discuss in Section 3.6.2 the need and how to introduce asymmetry into the along-strike variation of the μ parameter.

3.2.2 Model1.NO: Mixed-Effect Regression of Tier-1 Data

The first three refinements described above are implemented in the interim update *Model1.NO*. To handle data non-normality noted on Figure 3.2, we remove a limited number of small-displacement data to help bring the remaining data closer to normality. Identification of such small-displacement data is discussed in Section 3.2.2.1. Note that data trimming (removal of extreme values) is not applied to the right tail of data distribution (data with larger displacements), which matters to engineering application.

3.2.2.1 Tier Classification of Displacement Data

Data removal is a common practice and has been used extensively in data analysis where outliers are present. To identify data to be removed, we devise a methodology that utilizes the results of parametric and non-parametric quantile regressions (Koenker, 2005) to group displacement data into two tiers, on an earthquake-by-earthquake basis. Details about the methodology is given Appendix D.

Tier classification of a displacement point is based on its value. For most earthquakes and most locations on the main trace, the boundary between tier-1 and tier-2 classes corresponds to a small percentile of the available data values. Averaging over all earthquakes and all locations, the tier boundary corresponds to roughly the 15th percentile. We identify 244 data points (out of 3,334,

or roughly 7.5% of our dataset) as tier 2. A large portion of tier-2 data come from just a few earthquakes, most of which are relatively well-sampled earthquakes such as Landers, Hector Mine, and Izmit-Kocaeli. Tier classification of displacement data are shown on Figures 2.10 and 3.4 and also provided in column “Tier” of the electronic supplement 1.

3.2.2.2 Model for μ

Normal distribution is again assumed in *Modell.NO*, but formulation for its μ (mean) parameter is revised as follows,

$$\begin{aligned}
 \ln(D_{ij}) &\sim N(\mu_{ij}, \sigma^2) \\
 \mu_{ij}(\mathbf{M}_i, x_{ij}^*) &= c_0 + \delta_{eq,i} + f_M + f_{l2L} \\
 \delta_{eq,i} &\sim N(0, \sigma_{eq}^2) \\
 f_M &= m_2(\mathbf{M}_i - m_3) + \frac{m_2 - m_1}{c_n} \ln\left(\frac{1 + e^{-c_n(\mathbf{M}_i - m_3)}}{2}\right) \\
 f_{l2L} &= c_1(x_{ij}^* - 1)
 \end{aligned} \tag{3.2}$$

Equation (3.2) contains two improvements to P11’s original μ formulation in Equation (3.1):

- A random effect on the intercept of the μ model, $c_0 + \delta_{eq,i}$
 - Random intercept $\delta_{eq,i}$ represents the random earthquake effect on the mean parameter μ .
 - Three types of data residuals are computed for use in the diagnosis of a random-intercept GMM, and we also use them in the diagnosis of a random-intercept FDM assuming normality, such as the *Modell.NO* discussed here. Residual with respect to μ_{ij} is called the within-earthquake residual; $\delta_{eq,i}$ is called the between-earthquake residual; and residual with respect to $(c_0 + f_M + f_{l2L})$ is called the total residual. The variance of these types of residuals is called the within-earthquake variance (σ^2), between-earthquake variance (σ_{eq}^2), and total variance, respectively.
- A bilinear \mathbf{M} scaling relation, f_M
 - The \mathbf{M} -scaling term f_M in Equation (3.2) is taken from the GMM of Chiou and Youngs (2008, 2014). Its coefficients may be interpreted as follows. Coefficients m_1 and m_2 represent roughly the linear \mathbf{M} -scaling rate in the range of $\mathbf{M} \ll m_3$ and $\mathbf{M} \gg m_3$, respectively. A transition between m_1 and m_2 occurs in a magnitude range centered at $\mathbf{M} = m_3$; for this reason, m_3 is named the hinge magnitude of

bilinear \mathbf{M} -scaling. The sharpness of the transition is determined by the value of coefficient c_n ; the larger the c_n , the sharper the transition.

- Function f_M is nonlinear in coefficients m_3 and c_n .
- The denominator 2 inside the ln function is not in the original f_M of Chiou and Youngs (2008, 2014); it is included here so that c_0 can be interpreted as the mean parameter for $\mathbf{M} = m_3$ at location $l2L = 0.5$; more discussions are given below.

In addition, the $c_1 x_{ij}^*$ term of Equation (3.1) is revised to $c_1 (x_{ij}^* - 1)$. Centering to $x^* = 1$ (or $l2L = 0.5$) does not change the estimated c_1 , but it allows the interpretation of coefficient c_0 as the mean displacement at the midpoint of an $\mathbf{M} = m_3$ rupture trace. In contrast, P11's c_0 is the mean at the edge of the main trace of an $\mathbf{M} = 0$ rupture, an un-observable surface rupture scenario. Note that P11 and the FDMs discussed in this report do not ensure that predicted D is near zero at both edges of the main trace by tapering μ to a sufficiently negative value.

Geometrically, $c_0 + c_1 (x^* - 1)$ tracks the upper half of an ellipse centered at ($l2L = 0.5$, c_0). One can relax the assumption of along-strike symmetry by horizontally shifting the center away from $l2L = 0.5$ or adding an extra linear $l2L$ term to the μ model (see Section 3.6.2).

Estimation of *Modell.NO* is carried out using the nonlinear mixed-effect regression method (Pinheiro and Bates, 2000), and the estimated coefficients are listed in Table 3.1. A nonlinear regression is needed to estimate the nonlinear coefficients m_3 and c_n in f_M . Because the available earthquakes do not span a sufficiently large magnitude range, coefficient c_n is not precisely estimated; the estimated value is about 232, and the estimation standard error is enormous (about 128,000). Having such a large c_n amounts to a pieces-wise linear relation, which we do not favor. We judge that $c_n = 10$ is a proper choice because it provides an equally good fit to the data as $c_n = 232$ does, yet it renders a reasonably gradual transition of scaling rate from m_1 to m_2 . Coefficient m_3 is estimated via the nonlinear mixed-effect regression.

3.2.2.3 Model for Within-Earthquake Variance

The improvement on the within-residual variance (σ^2) as discussed in Section 3.2.1.2 is implemented here. It is reasonable to assume mid-point symmetry in both μ and σ^2 . To ensure that σ^2 is symmetric about $l2L = 0.5$, we choose $l2L_f$ as the covariate of the σ^2 model.

To help study the along-trace variation of σ^2 , we conduct a trial mixed-effect regression on tier-1 data, assuming homoscedasticity (σ^2 is a constant). The within-earthquake residuals from this trial run are then plotted on Figure 3.5. As a visual aid to help delineate the along-trace variation, standard deviations of within-earthquake residuals in 10 non-overlapping $l2L_f$ bins are computed and plotted on Figure 3.5 after multiplying them by 1.65. This figure clearly indicates that σ^2 decreases with increasing $l2L_f$. This downward trend has been observed in previous studies. Youngs et al. (2003) stated that combining the displacement data symmetrically about $l2L = 0.5$ can result in a larger variability that reflects the asymmetric pattern of measured

displacement. P11 reported, in their bilinear model, a higher residual standard deviation for locations near the rupture ends ($l2L_f < 0.30$) than for locations near the center.

The observed $l2L_f$ dependence of binned σ^2 on Figure 3.5 can be adequately modeled by either a linear or an exponential function of $l2L_f$. These two functions cannot be differentiated by data in the tight $[0, 0.5]$ range of $l2L_f$. We choose to use the latter, mainly because the logarithmic function is the default link function for the scale parameter of generalized linear model. Hence, we model the within-earthquake variance as follows:

$$\sigma^2 = c_{c_{v3}}^2 (e^{c_{v4} * l2L_f})^2 \quad (3.3)$$

The p -value of the likelihood ratio between the trial model (a constant σ^2 , equivalent to $c_{v4} = 0$) and *Modell.No* (the σ^2 model specified by Equation (3.3) and listed in Table 3.1) is less than 0.0001, which strongly rejects the assumption of homoscedasticity.

3.2.2.4 Variance Model of Random Intercept

In GMMs, the variance of random intercept has been found to be a function of magnitude (Youngs et al., 1995). To study the \mathbf{M} -dependence of σ_{eq} in FDM, the estimated $\delta_{eq,i}, i = 1, N_{eq}$, where $N_{eq} = 29$, of *Modell.NO* are plotted on Figure 3.6. As a visual aid, the standard deviations of $\delta_{eq,i}$ in five non-overlapping \mathbf{M} bins are also plotted on Figure 3.6. From this figure, σ_{eq} clearly decreases with increasing \mathbf{M} . We choose to model the observed trend by an exponential function of \mathbf{M} . However, no small earthquakes are in the dataset to validate the large σ_{eq} in the $\mathbf{M} < 6$ range that would have been predicted by extrapolating the fitted trend. Furthermore, the mechanical reason for such dependence on \mathbf{M} is not yet understood, and a theoretical justification supporting the large σ_{eq} value at $\mathbf{M} < 6$ is not known to us. Hence, we decide to impose a cap on the value of σ_{eq} and model it as follows:

$$\sigma_{eq} = c_{v1} e^{c_{v2} * \max(\mathbf{M}-6.1, 0)} \quad (3.4)$$

The cap is set to the fitted σ_{eq} at $\mathbf{M} = 6.1$, near the lower magnitude bound of selected strike-slip earthquakes.

Estimation of coefficients (c_{v1}, c_{v2}) is conducted outside the mixed-effect regression by regressing Equation (3.4) on $\delta_{eq,i}$ using a generalized nonlinear least-squares method (Pinheiro and Bates, 2000) in which σ_{eq}^2 can be modeled as a function of covariates. The likelihood ratio statistics against the constant- σ_{eq}^2 null hypothesis (equivalent to $c_{v2} = 0$) has a p -value of 0.02, indicating the null hypothesis can be rejected at the significance level of 0.05. The estimated c_{v1} and c_{v2} of *Modell.NO* are tabulated in Table 3.1.

The number of earthquakes in the $\mathbf{M} > 7.5$ range is insufficient to justify an unceasing downward trend of σ_{eq} toward 0. Therefore, for forward application, we further impose a floor on the value of σ_{eq} (a minimum σ_{eq}). We judge that 0.4, which is roughly the σ_{eq} in the highest magnitude bin on Figure 3.6, is a reasonable floor on σ_{eq} . Hence, the model for σ_{eq} is revised to

$$\sigma_{eq} = \max(c_{v1} e^{c_{v2} * \max(\mathbf{M}-6.1, 0)}, 0.4) \quad (3.5)$$

We note that, in the ground-motion model of Chiou and Youngs (2008, 2014), between-earthquake standard deviation (denoted as τ in their papers) is also capped on the low end of \mathbf{M} and floored on the high end of \mathbf{M} .

3.2.2.5 Predictive Distribution of Model1.NO

The predictive distribution of the response variable $\ln(D)$, $\ln(D | \delta_{eq}, \mu, \sigma^2) \sim N(\delta_{eq} + \mu, \sigma^2)$, is a compound distribution because the location (mean) parameter $\delta_{eq} + \mu$ is itself a random variable taken from the normal distribution $N(\mu, \sigma_{eq}^2)$. In forward application, if the value of δ_{eq} cannot be pre-specified, we should treat δ_{eq} as a source of aleatory variability. As such, the full range of δ_{eq} value should be considered when evaluating the predictive distribution of $\ln(D)$, by either analytically or numerically integrating out δ_{eq} while taking into account the probability density of δ_{eq} . The integration (or marginalization) over $\delta_{eq} \sim N(0, \sigma_{eq}^2)$ leads to a distribution no longer conditional on δ_{eq} . Note that these descriptions about the predictive distribution also apply to other random-intercept FDMs, where the normal distribution of response variable is replaced with a distribution pertinent to the particular FDM.

It is well known that the distribution resulting from compounding the normal distribution $N(m, \sigma^2)$ with the normal distribution $N(\mu, \sigma_{eq}^2)$ of m leads to another normal distribution $N(\mu, (\sigma')^2)$, where

$$(\sigma')^2 = \sigma^2 + \sigma_{eq}^2 \quad (3.6)$$

Another interpretation of compounding is that $\ln(D)$ is a continuous mixture of normal random variable $y \sim N(m, \sigma^2)$ for which the mean parameter m is normally distributed.

The 0.05, 0.50, and 0.95 quantiles predicted by the compound distribution from *Model1.NO* are shown on Figure 3.7. The 0.95 quantile is smaller than P11's, except for \mathbf{M} around 7 and for $l2L \leq 0.2$. Relative to P11, the adopted bilinear \mathbf{M} relationship yields a steeper magnitude scaling for $\mathbf{M} < 7.1$ (m_1 is larger than the reported scaling rate of 1.7927 in P11) but a shallower scaling for $\mathbf{M} > 7.1$ ($m_2 < 1.7927$). The latter, together with the \mathbf{M} -dependent σ_{eq}^2 , render, in our opinion, a more reasonable model extrapolation for $\mathbf{M} > 8$ that results in a 0.95 quantile more commensurate with the observed D_{max} of past large earthquakes discussed in Section 3.2.1.2.

In addition to the sharp break near $\mathbf{M} = 7.1$ (the hinge magnitude of bilinear \mathbf{M} scaling relation for μ), the \mathbf{M} -dependent σ_{eq} causes two gentle breaks in slope near $\mathbf{M} = 6.1$ (reaching the floor on σ_{eq}) and $\mathbf{M} = 7.5$ (reaching the cap on σ_{eq}). The along-trace variation (bottom panel of Figure 3.7) deviates slightly from an ellipse because of the $l2L_f$ -dependence in σ^2 .

The differences between *Model1.NO* and P11 noted in the last two paragraphs are due to the difference in regression data as well as the three formulation refinements implemented in *Model1.NO*. Similar differences are also noted for the other updates to be discussed later. To avoid

redundancy and save space, these observed differences will not be repeated in the following sections. All updates are summarily compared in Section 3.4.

3.2.2.6 Residual Diagnostics

Diagnostic plots of the residuals of *Modell.NO* are presented on Figure 3.8. Compared to those of the simple update on Figure 3.2, the standardized within-earthquake residuals from *Modell.NO* are closer to a standard normal distribution. However, the fit still has room for improvement; the residual histogram indicates a left-skewed distribution with a right tail thinner than that of the standard normal distribution. These deficiencies indicate that data trimming, as helpful as it can be, is not capable of turning non-normal data into a normal dataset.

3.2.3 Model2.SN: Distributional Regression Using Skew-Normal Distribution

3.2.3.1 Skew-Normal Distribution

Our second effort of handling data non-normality is to replace normal distribution by the three-parameter SN distribution. SN distribution (Azzalini and Capitanio, 2014; Rigby et al., 2020, page 378) is a generalization of normal distribution to introduce asymmetry to the probability density function. The parameters of SN distribution are denoted as μ (location parameter, real value), σ (scale parameter, positive real value), and ν (skew parameter, real value). The extra parameter ν regulates jointly the skewness and the kurtosis of the SN distribution. When ν approaches 0, SN distribution specializes to the normal distribution. A negative ν yields a left-skewed distribution, and a positive ν yields a right-skewed distribution. As ν approaches $+\infty$, SN approaches the (positive) half normal distribution.

For simplicity, we use the same symbols for the distribution parameters of different types of distribution, but note that they have different interpretations, and they play different roles under different distributions. For example, μ and σ are the mean and the standard deviation of normal distribution, respectively. But in SN distribution, μ is not the mean, and σ is not the standard deviation.

The basic characteristics of the SN distribution (Rigby et al., 2020, Table 18.9 on page 379) are related to the distribution parameters in the following ways:

- Mean = $\mu + \sqrt{\frac{2}{\pi}} \sigma \frac{\nu}{\sqrt{1+\nu^2}}$,
- Standard deviation = $\sigma \sqrt{1 - \frac{2}{\pi} \frac{\nu^2}{(1+\nu^2)}}$,
- Skewness = $\text{sign}(\nu) \frac{4-\pi}{2} \left[\frac{\pi}{2} (1 + \nu^{-2}) - 1 \right]^{\frac{3}{2}}$,
- Excess kurtosis = $2(\pi - 3) \left[\frac{\pi}{2} (1 + \nu^{-2}) - 1 \right]^{-2}$

As indicated in the above expressions, parameter ν regulates not only the skewness and the kurtosis, but also the mean and the standard deviation. In contrast, location parameter μ regulates only the mean. Scale parameter σ regulates both the mean and the standard deviation.

The probability density function (PDF) and the cumulative distribution function (CDF) from Rigby et al. (2020, Table 18.9 on page 379) are reproduced in Appendix C.1. These two functions are implemented in FORTRAN, and the source code is available from <https://www.conservation.ca.gov/cgs/pfdha>. Examples of the PDF and the CDF of SN distribution are shown in Section 3.4.

3.2.3.2 GAMLSS Regression

The mixed-effect model and estimation method used in the development of *Modell.NO* are not viable here because the underlying distribution is no longer the normal distribution. From this point on, we use the generalized additive model for location, scale, and shape (GAMLSS), whose estimation is carried out using R's *gamlss* package (Stasinopoulos et al., 2017).

GAMLSS is an extension of the generalized additive model (Hastie and Tibshirani, 1990; Wood, 2017), which itself is an extension of the generalized linear model (McCullagh and Nelder 1989). GAMLSS targets the complete distribution of the response variable, not just its mean parameter, as is the case of traditional linear model. The relations between the distribution parameter and a set of explanatory variables can be modeled in GAMLSS. As such, GAMLSS is often regarded as a distributional regression model (Kneib et al., 2021). The basic GAMLSS specification includes the following:

- Probabilistic distribution of the response variable
 - GAMLSS can model distributions outside the exponential family of distribution, including the skew-normal, the skew- t , and the nEMG distributions used in this study.
- For each distribution parameter, a link function that maps the nonlinear relation between a distribution parameter and its explanatory variables to a linear relation
 - In this study, log-link function is used for a positive-valued parameter (such as the scale parameter σ of the SN distribution), and identity link for a real-valued parameter (such as the μ and the ν parameters of SN distribution).
- The linear relation between the link-transformed distribution parameter (predictor, in GAMLSS parlance) and the explanatory variables
 - In our study, the explanatory variables are **M** and *l2L*. We have not explored other covariates (such as soil condition and rupture complexity) that may have explanatory/predictive power on the distribution parameters.

- The linear relation may include parametric terms as well as nonparametric smoothing functions.
- A parametric term may include random effects on intercept or slopes
 - The linear mixed-effect regression model described in Section 3.2.3.1 can be regarded as a member of GAMLSS.

Although random effects can be on both intercept and slopes in any of the linear relation for the distribution parameter, we allow only random intercept in the μ model. This restriction is necessitated by the difficulty of analytically marginalizing over random slopes. Analytic expression for distribution marginalized only over a normally distributed random intercept can be derived more easily, avoiding numerical integration altogether. We do recognize the scientific merits of having random effects on the scale and shape parameters. The resulting information, for example, may provide insights into potential relations between the complexity of a rupture and the observed level of variability in displacement. Such undertaking is not in the scope of this study.

The *gamlss* package provides three types of additive terms for modeling and estimating the random intercept of a GAMLSS. We prefer the *gamlss.re* function because it is built on the *nlme* package also used by Chiou and Youngs (2008, 2014) in GMM development. One limitation of *gamlss*, however, is that it supports only linear relations, while our model for μ is nonlinear in coefficient m_3 . Such limitation of the *gamlss* package is circumvented in this study by fixing the nonlinear coefficient m_3 to a specific credible value so that the resulting **M**-scaling relation is a linear function of the remaining coefficients; more discussions on the credible range of m_3 and the accompanying prediction uncertainty are given in Section 3.2.3.3 and Section 3.5, respectively.

Even though smoothing functions (such as loess and penalized smoothing spline) may be included, our model formulation is purely parametric, without smoothing terms. However, we did use smoothing functions to guide the development of the functional forms of some distribution parameters. The outcomes are consistent with the findings through data binning; therefore, to save space, we omit presentations on the exploratory study of functional forms using smoothing functions.

The above paragraphs provide a short description of the structure of the GAMLSS used in our study. More information can be found in Stasinopoulos et al. (2017). The descriptions given here apply to all GAMLSS models covered in this Chapter.

3.2.3.3 Models for μ , σ , and ν

Model of parameter μ is the same as that of *Modell.NO* given in Section 3.2.2.2 (Equation (3.2)). Here, we use the representative m_3 value of 7.1 for the purpose of describing the linearized model formulation for μ . The linearized μ model is

$$\mu_{ij}(\mathbf{M}_i, x_{ij}^*) = c_0 + \delta_{eq_i} + m_2(\mathbf{M}_i - 7.1) + \frac{(m_2 - m_1)}{c_n} \ln\left(\frac{1 + e^{-c_n(\mathbf{M}_i - 7.1)}}{2}\right) + c_1(x_{ij}^* - 1) \quad (3.7)$$

Definitions of coefficients are given earlier below Equation (3.2). Coefficient c_n is again set to 10.

For the scale parameter σ , we use the log-link function and a linear predictor model with the explanatory variable $l2L_f$

$$\ln(\sigma_{ij}) = \ln(c_{v3}) + c_{v4} * l2L_{f,ij}$$

or

$$\sigma_{ij} = c_{v3} e^{c_{v4} * l2L_{f,ij}} \quad (3.8)$$

Equation (3.8) has the same algebraic expression as that of the within-earthquake variance model of *Modell.NO* (Equation (3.4)), but it models the scale parameter, not the variance of SN random variable.

The skew parameter v is modeled as a constant (that is, same for every observation),

$$v_{ij} = c_{v5} \quad (3.9)$$

For the purpose of modeling the along-trace variation of variance, we could have introduced $l2L_f$ dependence into the v model, instead of the σ model. We decide not to do so because v regulates all four basic characteristics of the SN distribution; thus, having an $l2L_f$ -dependent v would have incurred an $l2L_f$ -dependent mean (which already depends on $l2L_f$ via the μ model), skewness, and excess kurtosis. To keep it simple, we prefer having constant skewness and excess kurtosis, and thus the decision was not to include $l2L_f$ dependence in the v model.

3.2.3.4 Variance Model of Random Intercept

As in *Modell.NO*, we use Equation (3.4) to model the variance σ_{eq}^2 of $(\delta_{eq,i}, i = 1, N_{eq})$ and regress for the two coefficients using the generalized nonlinear least-squares method. For forward application, we also recommend using Equation (3.5) with a floor of 0.4 on σ_{eq} . The recommended model is shown on Figure 3.9. The estimated coefficients (c_{v1}, c_{v2}) are tabulated in Table 3.1

3.2.3.5 Diagnostics of Model2.SN Using Normalized Quantile Residual

The residual defined in Section 3.2.1.1, and the within-earthquake residuals defined in Section 3.2.2.2 are useful for regression model that assumes a normal distribution for the response variable. Stasinopoulos et al. (2017, Section 12.2) suggest using the normalized quantile residual \hat{r}_{ij} (Dunn

and Smyth, 1996) for checking a GAMLSS model where a non-normal distribution is assumed for the response variable.

Normalized quantile residual is computed as follows. First, cumulative probability of an observed $\ln(D_{ij})$ value is computed using the estimated distribution parameters for that observation. The computed probability is then mapped to a standard normal variate through the quantile function (inverse of cumulative distribution function) of standard normal distribution. The normalized quantile residual of $\ln(D_{ij})$ is taken as the value of mapped standard normal variate and denoted by \hat{r}_{ij} . Expressed in a compact form, the normalized quantile residual is defined as

$$\hat{r}_{ij} = \Phi^{-1}[F(\ln(D_{ij}) | \hat{\theta}_{ij})] \quad (3.10)$$

where $\hat{\theta}_{ij}$ is a vector containing the estimated distribution parameters for observation (i, j) ; $F(\ln(D_{ij}) | \hat{\theta}_{ij})$ is the cumulative probability function of $\ln(D_{ij})$ given $\hat{\theta}_{ij}$; and $\Phi^{-1}[\cdot]$ is the inverse CDF of the standard normal distribution. The definition of \hat{r}_{ij} is general and applicable to other continuous distributions to be discussed later. If the estimated distributional model is correct, \hat{r}_{ij} will follow the standard normal distribution, aside from sampling variability.

Five diagnostic plots of the normalized quantile residuals are presented on Figure 3.10. The histogram and the Q-Q plot of \hat{r}_{ij} both reveal a slight deviation from the standard normal distribution, indicating that SN is a much better distribution than the normal distribution (as used in the simple update and *Model1.NO*) in characterizing the probabilistic distribution of tier-1 data. These observations also indicate that *Model2.SN* is an adequate model and additional flexibility is not crucially needed for tier-1 data. However, when we bring the tier-2 data back into regression analysis, the SN-based model struggles to fit them. Our aspiration to have a complete model for both tiers of data leads to the third update described below in Section 3.2.4.

3.2.3.6 Predictive Distribution by Model2.SN

Analogous to *Model1.NO*, $\ln(D)$ predicted by *Model2.SN* follows the distribution that results from compounding $SN(m, \sigma, \nu)$ with the normal distribution $N(\mu, \sigma_{eq}^2)$ of m . This compounding may also be interpreted as a continuous mixture of SN random variable for which the location parameter is normally distributed. The compound distribution is the skew-normal distribution $SN(\mu, \sigma', \nu')$ (Arellano-Valle et al., 2005), where

$$\begin{aligned} \sigma' &= \sqrt{\sigma^2 + \sigma_{eq}^2} \\ \nu' &= \frac{\sigma}{\sqrt{\sigma^2 + \sigma_{eq}^2}} \frac{\nu}{\sqrt{1 + \frac{\nu^2}{\sigma^2} \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_{eq}^2}\right)^{-1}}} \end{aligned} \quad (3.11)$$

Example quantiles predicted by the compound distribution of *Model2.SN* are shown on Figure 3.11. They are somewhat different from those by *Model1.NO* (Figure 3.7), mainly due to the difference in assumed response distribution. A more in-depth comparison of model predictions is given in Section 3.4.

3.2.4 *Model3.ST*: Distributional Regression Using Skew-*t* Distribution

3.2.4.1 *Skew-t* Distribution

In *Model3.ST*, the three-parameter skew-normal is replaced by the four-parameter ST distribution (Azzalini and Capitanio, 2014; Rigby et al., 2020, page 412); the extra parameter of ST distribution allows the proper modeling of both tiers of data. Although there are other types of probability distribution that also provide extra flexibility, we pick the skew-*t* distribution because it is a generalization of Student's *t* distribution through a mathematical manipulation similar to that used in generalizing the normal to the skew-normal. Incidentally, including the tier-2 data actually yields better fit compared to the skew-*t* fitted to only tier-1 data.

Parameters of the skew-*t* distribution (as parameterized in the *gamlss* package) are denoted as μ (location parameter, real value), σ (scale parameter, positive real value), ν (skew parameter, real value), and τ (excess kurtosis parameter, positive real value). Basic characteristics of skew-*t* distribution are related to the distribution parameters in complex ways. These relations are reproduced below from Rigby et al. (2020, Table 18.25, page 413),

- Mean = $\mu + \sigma \delta b_\nu$, where $\delta = \frac{\nu}{\sqrt{1+\nu^2}}$ and $b_\nu = \sqrt{\frac{\Gamma(\frac{\tau-1}{2})}{\Gamma(\frac{\tau}{2})} \frac{\tau}{\pi}}$
- Variance = $\sigma^2(\frac{\tau}{\tau-2} - \delta^2 b_\nu^2)$
- Skewness = $\sigma^3[\frac{\tau(3-\delta^2)}{\tau-3} \delta b_\nu - 3(\frac{\tau}{\tau-2} - \delta^2 b_\nu^2) \delta b_\nu - (\delta b_\nu)^3] / [\sigma^2(\frac{\tau}{\tau-2} - \delta^2 b_\nu^2)]^{1.5}$
- Kurtosis = $\sigma^4[\frac{3\tau^2}{(\tau-2)(\tau-4)} - 4\frac{\tau(3-\delta^2)}{\tau-3} \delta b_\nu + 6(\frac{\tau}{\tau-2} - \delta^2 b_\nu^2) \delta^2 b_\nu^2 + 3\delta^4 b_\nu^4] / [\sigma^2(\frac{\tau}{\tau-2} - \delta^2 b_\nu^2)]^2 - 3$

The PDF of the skew-*t* distribution are reproduced in Appendix C.2, and some examples are shown in Section 3.4. In addition to the implementations in R (in the *sn* and *gamlss.dist* packages), skew-*t*'s PDF is implemented by us using the FORTRAN language, and the CDF is computed via numerical integration of the PDF. The FORTRAN source code is available at <https://www.conservation.ca.gov/cgs/pfdha>.

3.2.4.2 Models for μ , σ , ν , and τ

Identity link is used for μ and ν , while log-link is used for σ and τ . The linear predictor models for μ , σ , and ν are the same as those of *Model2.SN*; hence, they are not repeated here. Parameter τ is modeled as a constant as follows:

$$\tau_{ij} = c_{v7} \quad (3.12)$$

Estimated coefficient values are tabulated in Table 3.1.

3.2.4.3 Variance Model of Random Intercept

As in *Model1.NO* and *Model2.SN*, standard deviation of $(\delta_{eq,i}, i = 1, N_{eq})$ is modeled as an exponential function of \mathbf{M} , with a cap on the low end of \mathbf{M} range and a floor on the high end of \mathbf{M} range (Equation (3.5)). The fitted σ_{eq} model and $\delta_{eq,i}$ values are shown on Figure 3.12, and estimated coefficients (c_{v1}, c_{v2}) are tabulated in Table 3.1.

3.2.4.4 Diagnostics of *Model3.ST*

Diagnostic plots using the normalized quantile residuals of *Model3.ST* are presented on Figure 3.13. These plots indicate that the ST distribution fits both tiers of displacement data quite well. Hence, we judge that further attempt to improve the current fit by introducing additional distribution parameters is superfluous.

3.2.4.5 Predictive Distribution of *Model3.ST*

Analogous to *Model1.NO* and *Model2.SN*, $\ln(D)$ predicted by *Model3.ST* follows the distribution that results from compounding $ST(m, \sigma, \nu, \tau)$ with the normal distribution $N(\mu, \sigma_{eq}^2)$ of the location parameter m . The predicted $\ln(D)$ can also be interpreted as a continuous mixture of ST random variable for which the location parameter of its distribution is normally distributed. We have not yet secured an analytic expression for the PDF of compound skew- t distribution. For now, we rely on stochastic simulation to compute approximately the quantiles of the compound distribution of $\ln(D)$. For a given rupture scenario (specified in terms of \mathbf{M} and $l2L$), we take 50 random samples of the mean parameter m from the distribution $N(\mu, \sigma_{eq}^2)$, and then 3,000 samples from the skew- t distribution $ST(m, \sigma, \nu, \tau)$ for each of the 50 sampled m . A quantile of the compound skew- t distribution is taken as the corresponding quantile of the 150,000 skew- t samples.

Examples of quantiles predicted by the compound *Model3.ST* are shown on Figure 3.14. Overall, we do not see any feature or trend about their differences from P11's predictions that we have not seen in *Model1.NO* and *Model2.SN*.

3.3 MODEL4.NEMG: DISTRIBUTIONAL REGRESSION USING NEGATIVE EXPONENTIALLY MODIFIED GAUSSIAN DISTRIBUTION

Model3.ST fits both tiers of displacement data well. We would have ended our model development and made *Model3.ST* our preferred model if not for the following shortcomings. Compared to the normal distribution, it is more difficult to interpret ST's distribution parameters, and it is more time-consuming to numerically compute the CDF (Appendix C.2) and to compound the (conditional) ST distribution with the normal distribution of the latent (unobservable) random intercept (Section 3.2.4.5). In addition, ST distribution lacks a geological underpinning. The nEMG distribution improves on the ST distribution in these regards. Development of the nEMG-based FDM is presented in this Section. This model, *Model4.nEMG*, is our preferred model for $\ln(D)$ among the models described in this report.

In this Section, we first propose a connection between the stochastic representation of nEMG random variable and the complexity of surface rupture (Section 3.3.1). We then discuss nEMG's probability distribution in Section 3.3.2. Regression relations for the three distribution parameters are presented in Section 3.3.3. Estimation of the model coefficients is carried out using R's *gamlss* package, and the results are given in Table 3.1. Estimated variance model of random intercept is presented in Section 3.3.4, and diagnostics of *Model4.nEMG* are presented in Section 3.3.5. Finally, quantiles predicted from *Model4.nEMG* are presented in Section 3.3.6.

3.3.1 Stochastic Representation of $\ln(D)$

It is shown above that normal distribution is inadequate in modeling the observed $\ln(D)$. Our search for a replacement distribution has so far been phenomenological, that is, not considering the mechanical process that generates the heavy left tail. Several studies (for examples, Scholz and Lawler, 2004; Martel and Shacat, 2006) suggested that interaction between adjacent segments in an earthquake results in tapering of surface displacement toward the interacting tip of each segment. We interpret such interior tapering as the main mechanism that generates the heavy left tail in the observed $\ln(D)$ distribution. Note that the exterior tapering toward the two rupture edges is already being modeled as part of the ellipse formulation for the location parameter. Based on the above, we propose that surface displacement in the absence of interior interaction effect be modeled by a normal variate G and the tapering effects of interior interaction be modeled by subtracting an exponential variate E from G ; G and E are assumed to be independent. We thus have the stochastic representation

$$\ln(D) = G - E \quad (3.13)$$

where $G \sim N(\mu, \sigma)$ and $E \sim Exp(\nu)$ are the independent normal and exponential components, respectively.

Geologically, the ν parameter value is expected to be highly variable along the main trace because displacement taper occurs sporadically. The frequency of tapers (Manighetti et al., 2015)

and their along-trace locations differ between earthquakes, making ν vary substantially between earthquakes. To keep FDM simple, we do not model these detailed complexities. Instead, we model the composite ν of the 29 selected strike-slip ruptures, which is expected to vary smoothly along the main trace. It should be noted that, via the composite ν , the taper's frequency and location variabilities are captured, to the extent present in the FDHI data, by the estimated distributional model; analysts should be mindful of potential double counting such variability in forward application.

Probabilistically, subtracting the exponential variate E helps transform the symmetric normal distribution to a left-skewed distribution, conforming to the observed left-skew distribution of $\ln(D)$.

3.3.2 nEMG Distribution

The $(G - E)$ random variate is not widely known, and to our knowledge, its probabilistic distribution has not been given a formal name. Its negative, $(-G + E)$, is better-known and has a right-skewed distribution that has been named the *exponentially modified Gaussian* (EMG, or *exGaussian*) distribution (Lovison and Schindler, 2014; Gori and Rioul, 2019; Rigby et al., 2020). Prompted by this relation, we call the distribution of $(G - E)$ the *negative exponentially modified Gaussian* distribution, or nEMG distribution for short. Its statistical characteristics, CDF, and PDF are readily derived from those of the better-known EMG distribution.

The three distribution parameters of the nEMG distribution are denoted by μ (the mean of Gaussian component, real value), σ (the standard deviation of Gaussian component, positive real value), and ν (the mean and the standard deviation of exponential component, positive real value). The commonly used statistical measures are related to the distribution parameters as follows:

- Mean = $\mu - \nu$,
- Variance = $\sigma^2 + \nu^2 = \nu^2 \left[1 + \frac{\sigma^2}{\nu^2}\right]$,
- Skewness = $-2 \left[1 + \frac{\sigma^2}{\nu^2}\right]^{-1.5}$,
- Excess kurtosis = $6 \left[1 + \frac{\sigma^2}{\nu^2}\right]^{-2}$

Compared to the SN and ST distributions described earlier, the above relations are remarkably simple. The skewness is always negative because the nEMG distribution is left-skewed. The skewness and excess kurtosis are regulated by the ratio $\frac{\sigma^2}{\nu^2}$. As ν approaches zero, the statistical measures of nEMG variate tend toward those of its normal component G . Larger ν yields a heavier left tail. The right tail of the nEMG distribution is not identical to the right tail of the normal distribution.

Analytic expressions of the nEMG distribution's PDF and CDF are given in Appendix C.3. A FORTRAN implementation of PDF and CDF is also included in Appendix C.3. Example PDF and CDF are shown in Section 3.4.

3.3.3 Models for μ , σ , and v

Model formulations for the three distribution parameters of the nEMG distribution are given as follows:

$$\begin{aligned}
\mu_{ij}(\mathbf{M}_i, x_{ij}^*) &= c_0 + \delta_{eq_i} + f_M(\mathbf{M}_i) + f_{l2L}(l2L_{ij}) \\
\delta_{eq_i} &\sim N(0, \sigma_{eq}^2) \\
f_M(\mathbf{M}_i) &= m_2(\mathbf{M}_i - m_3) + \frac{m_2 - m_1}{c_n} \ln\left(\frac{1 + e^{-c_n(\mathbf{M}_i - m_3)}}{2}\right) \\
f_{l2L}(l2L_{ij}) &= c_1(x^* - 1) \\
\sigma_{ij} &= c_{v3} \\
v_{ij} &= c_{v5} e^{c_{v6} * l2L_{f,ij}}
\end{aligned} \tag{3.14}$$

The model for μ is identical to that previously used in *Model2.SN* and *Model3.ST*; it is repeated here for completeness. As in previous models, coefficient c_n is set to 10. Coefficient m_3 is fixed to a credible value to linearize $f_M(\mathbf{M}_i)$. Here, we use $m_3 = 7.1$ as the representative value for the purpose of presentation. More discussions on the range of credible m_3 values are given in Section 3.5.1.

Along-main-trace variation of the variance of $\ln(D)$ can be included via either σ or v . Unlike *Model2.SN* and *Model3.ST*, we choose to model σ as a constant and model v as an exponential function of $l2L_{f,ij}$. This choice is rooted in the intuition that the Gaussian component represents a smooth (homogenous) rupture, hence a constant σ , whereas the exponential component reflects complexity associated with the interplay of interacting segments, hence an $l2L_f$ -dependent v . A consequence of this modeling decision is that the skewness and kurtosis of the predicted nEMG distribution will vary with $l2L$, and the predicted mean has two sources of along-main-trace variation, one from μ and one from v .

Without an existing implementation of the nEMG distribution in the *gamlss* package, our regression is carried out using the EMG distribution (function *exGAUS* in the *gamlss* package) on the negative $\ln(D)$. The regression results are then converted to those appropriate for $\ln(D)$ and tabulated in Table 3.1.

3.3.4 Variance Model of the Random Intercept of μ

Similar to the interim updates, the standard deviation σ_{eq} of $\delta_{eq,i}$ is modeled as an exponential function of \mathbf{M} , with a cap on the low end of \mathbf{M} range and a floor on the high end of \mathbf{M} range. The fitted σ_{eq} model and the $\delta_{eq,i}$ values are shown on Figure 3.15. Estimates of the σ_{eq} model coefficients are tabulated in Table 3.1. See Section 3.4 for a comparison of this σ_{eq} model with those from the interim updates.

3.3.5 Diagnostics of Model4.nEMG, Normalized Quantile Residual

Plots of the normalized quantile residuals of *Model4.nEMG* are shown on Figure 3.16. As mentioned earlier, when the estimated distributional model is appropriate, the model's normalized quantile residuals follow the standard normal distribution. The histogram, the empirical CDF, and the Q-Q plots on Figure 3.16 all indicate that normalized quantile residuals closely follow the standard normal distribution. Hence, we conclude that *Model4.nEMG* is an appropriate model providing good (distributional) fits to the displacement data.

3.3.6 Predictive Distribution

Analogous to other interim updates, $\ln(D)$ predicted by *Model4.nEMG* follows the distribution that results from compounding $nEMG(m, \sigma, \nu)$ with the normal distribution $N(\mu, \sigma_{eq}^2)$ of location parameter m . The predicted $\ln(D)$ can also be interpreted as a continuous mixture of the nEMG random variable for which the location parameter of its distribution is normally distributed. It is straightforward to show that the compound distribution is again an nEMG distribution $nEMG(\mu, \sigma', \nu)$ where $(\sigma')^2 = \sigma^2 + \sigma_{eq}^2$.

Quantiles predicted by the compound distribution of *Model4.nEMG* are shown on Figure 3.17. We notice large differences from the quantiles of P11, and these differences are generally consistent with those observed in the interim updates. Despite being redundant of earlier Sections, we again outline the observed differences below. The 0.95 quantile is much lower than P11's, except for \mathbf{M} near 7 and for $l2L_f \leq 0.2$. Compared to P11, the bilinear \mathbf{M} scaling of *Model4.nEMG* has a steeper scaling for $\mathbf{M} < 7.1$ (larger m_1) and a gentler scaling for $\mathbf{M} > 7.1$ (smaller m_2); these differences in scaling rates are due to the different \mathbf{M} -scaling formulations, regression methods, and regression datasets. The much smaller m_2 , together with the decreasing σ_{eq}^2 with \mathbf{M} , result in a much lower 0.95 quantile prediction for $\mathbf{M} > 8$. The \mathbf{M} dependence of σ_{eq} (which is not modeled in P11) is also responsible for two extra breaks in \mathbf{M} -scaling slope near $\mathbf{M} = 6.1$ (floor on σ_{eq}) and $\mathbf{M} = 7.4$ (cap on σ_{eq}), in addition to the sharp break near $\mathbf{M} = 7.1$ (the hinge magnitude m_3 of the bilinear \mathbf{M} scaling relation). Finally, the predicted $\ln(D)$ profile is flatter than P11's (smaller c_1). This difference in profile's flatness appears to be related to slip amplitude differences between the two datasets near the rupture edges, not to model formulation.

3.4 COMPARISON OF INTERIRM AND PREFERRED MODELS

So far five FDMs are developed using different combinations of tiered data and different distributions of response variable. A comparison of these models is presented in this Section to highlight their similarities as well as noticeable differences.

Table 3.1 provides a direct comparison of model coefficients of the μ parameter. Estimates of coefficients m_1 , m_2 , and c_1 are similar among models that assume non-normal distributions.

The estimated \mathbf{M} -dependence of σ_{eq} (coefficient c_{v2}) is similar among the models in Table 3.1, despite their differences in distribution assumption and data tier selection. The simple update (like in P11) does not include \mathbf{M} -dependent σ_{eq} in model formulation. For the purpose of comparison, we provide an estimate of c_{v2} for the simple update by regression on the average residual of individual earthquake (as defined in Section 3.2.1.4). The estimated c_{v2} for the simple update is less negative than the other models in Table 3.1. This outcome may be partially attributed to the use of linear \mathbf{M} -scaling relation in the simple update, which results in a poorer fit to data and a larger scatter in the average residuals in the upper magnitude range.

Direct comparison of the coefficients of σ and v parameters is not meaningful because their roles differ among the different types of distribution. Comparing the PDF and the CDF of the compound distributions resulting from the various FDMs is more informative. Such comparison is made for several \mathbf{M} and $l2L = 0.2$ on Figure 3.18. Another comparison of the spread and shape of compound distributions is shown on Figure 3.19, where the predicted 5th, 50th, and 95th percentiles are plotted in dot charts for several \mathbf{M} and several $l2L$ values. The interim updates and the final model are almost identical in formulation; the main difference between them is in the choice of response variable distribution. Hence, it is not unexpected that the predicted percentiles are similar among them. The larger model-to-model variation in the 5th percentile reflects the different shape of the left tail of assumed distributions.

From Figure 3.18, for the same \mathbf{M} , the skewness of the compound distribution from *Model4.nEMG* is less negative at $l2L = 0.5$ than at $l2L = 0$, a feature resulting from the $l2L_f$ -dependence in v , which affects the skewness. Also, for the same $l2L$, skewness is less negative at $\mathbf{M} 5.8$ (that is, distribution is more symmetric) than at $\mathbf{M} 8$, a feature due to the \mathbf{M} -dependent σ' of the compound distribution, which in turn induces \mathbf{M} -dependence in skewness. These features are manifested on Figure 3.19, which shows that the (5th – 50th) interpercentile is wider than the (95th – 50th) interpercentile. Similar features in skewness are also noted for *Model2.SN* and *Model3.ST*, but it is not as straightforward to pinpoint their causes. This is an advantage of the nEMG distribution over the other two distributions.

On Figure 3.20, we compare the predicted quantiles from six FDMs as a function of \mathbf{M} at several site locations. Quantiles by P11 and the simple update are quite different from those by the three interim updates and the preferred model (*Model4.nEMG*). These differences are mainly due

to the improved formulations described in Section 3.2. Compared to model formulation, predicted percentiles are less sensitive to the choice of probability distribution.

In Tables 3.2 to 3.5, we tabulate the values of the predicted distribution parameters as well as five different percentiles of the predicted distributions, for four magnitudes and two along-trace locations. This is mainly for the convenience of users who implement our FDMs and would like to verify their implementations.

Although the three interim updates are valid FDMs for $\ln(D)$, *Model4.nEMG* is our preferred model for the following reasons:

- Intuitive connection to the rupture mechanism responsible for the left-skewed distribution of $\ln(D)$ (Section 3.3.1)
- Simple interpretation of distribution parameters (Section 3.3.2)
- Exceptionally good fit to $\ln(D)$ data (Section 3.3.3)
- Ease of implementation afforded by the analytic expressions of CDF and PDF for both the conditional and the compound distributions of the predicted $\ln(D)$.

3.5 ASSESSMENT OF ESTIMATION UNCERTAINTY

In this section, we assess the estimation uncertainty of the distribution parameters of *Model4.nEMG*. We also evaluate the effects of estimation uncertainty on the predicted quantiles. Assessment presented in this report is focused on the estimation uncertainty of **M**-scaling relation and the accompanying prediction uncertainty.

3.5.1 Uncertainty in M-Scaling

To abide by the restriction of using linear parametric relation in GAMLSS, we fix the value of m_3 (the hinge magnitude around which the change of scaling rate takes place) to linearize the non-linear magnitude-scaling relation of the μ parameter. The resulting coefficient estimates are thus conditional on m_3 . Based on the analysis of *Modell.NO*, we learn that m_3 is negatively correlated with m_2 , whereas m_2 is critical to model's extrapolation outside data's upper **M** bound. Similarly, m_3 is negatively correlated with m_1 , whereas m_1 is critical to the extrapolation below data's lower **M** bound. Thus, it is important to assess the credible range of m_3 and the sensitivity of predicted quantiles to the selected m_3 .

In the following, we first assess the credible range of m_3 , then conduct regression using each of the four m_3 values selected to span the assessed credible range. Prediction uncertainty is then evaluated using the four resulting models. We judge that, if needed, these models can be used in different branches of a PFDHA logic tree to represent the epistemic prediction uncertainty related to the estimation uncertainty in **M**-scaling.

3.5.1.1 Credible range of m_3

The credible range of m_3 is assessed using estimation results of two sets of regression runs. The first set consists of only *Modell.NO*, where m_3 is estimated by the nonlinear mixed-effect regression on tier-1 data, assuming normality. The estimated m_3 of *Modell.NO* is 7.0621 and the reported standard error is 0.383.

The second set consists of multiple exploratory regression runs. Each run uses a GAMLSS whose μ parameter is without the \mathbf{M} -scaling term,

$$\mu_{ij}(x_{ij}^*) = c_{eq,i} + c_1(x_{ij}^* - 1) \quad (3.15)$$

As such, the earthquake-specific random coefficient $c_{eq,i}$ subsume the fixed-effect of \mathbf{M} as well as the earthquake-specific random effect $\delta_{eq,i}$. We then fit the bilinear relation $c_0 + f_M(\mathbf{M})$ to $c_{eq,i}$ using a least-squares regression to obtain an estimate of m_3 . Because the second set of regression runs is used mainly to guide the selection of the credible m_3 range, we ignore the estimation uncertainty of $c_{eq,i}$ in the least-squares regression. The exploratory regression is repeated for several combinations of probability distribution assumption and tiered data. The outcomes indicate that the m_3 estimates are remarkably stable, within the range of 7.07 to 7.09. The reported standard error of m_3 estimator is also remarkably stable, in the range of about 0.351 to 0.383. Regression analysis of $\ln(D_{ave})$ and $\ln(D_{max})$, to be presented in Section 3.6.3, result in m_3 estimates and standard errors that are also within or close to the above ranges.

Based on the above results and assuming m_3 is a normal variate, we assess that its mean is 7.1 (rounded up to the first decimal point for ease of memory) and its standard deviation is 0.35. This statistical assessment is subject further to the physical constraint discussed next. We find that the m_2 estimate is negative in value when the given m_3 is sufficiently large. As in GMM, we disallow oversaturation (negative m_2 leads to reversed scaling of $\ln(D)$ with \mathbf{M} for $\mathbf{M} > m_3$). Furthermore, negative \mathbf{M} -scaling is not supported by dynamic rupture modeling (Wang and Goulet, 2022). The constraint of a positive m_2 demands a right truncation on the assessed normal distribution for m_3 . We find that m_2 estimate is close to zero when $m_3 = 7.32$. We thus recommend using $m_3 = 7.32$ as the right truncation point.

As m_3 approaches 6 (the lower \mathbf{M} bound of our data) and falls below 6, estimation of m_1 becomes less reliable, if not infeasible. In addition, as m_3 decreases, m_2 estimate increases above 1.7927 (the \mathbf{M} -scaling coefficient reported by P11) and renders an unreasonably large 95th percentile prediction that we have worked hard to eradicate. We thus conclude that a left truncation on m_3 distribution is warranted, but we have insufficient data to assess the left truncation point. Instead, as a general guideline, we caution against using a m_3 value smaller than 6.4, which is the mean minus two standard deviations of the untruncated m_3 distribution. There are two bases to support this guideline. First, the cumulative probability of $m_3 = 6.4$ is already small (~ 0.0227).

Secondly, the accompanying m_2 will result in large 95th percentile predictions for $\mathbf{M} > 8.3$, which are hard to defend against the observed maximum displacements from currently available large- \mathbf{M} observations (Section 3.2.1.2) and the 10 to 20 m range of maximum simulated displacement from the dynamic rupture modeling of the \mathbf{M} 8 to \mathbf{M} 8.2 scenarios (Wang and Goulet, 2022).

3.5.1.2 Prediction uncertainty induced by uncertainty in m_3

We use four m_3 values to illuminate the prediction uncertainty induced by the epistemic uncertainty in m_3 . These four values are $m_3 = 6.4, 6.75, 7.10,$ and 7.32 , representing the mean minus two standard deviations, the mean minus one standard deviation, the mean, and the right truncation point of assessed m_3 distribution. For reference, we also include the over-saturation case of $m_3 = 7.45$. We conduct regressions to estimate the remaining model coefficients, conditional on a particular m_3 value. The resulting \mathbf{M} -scaling curves are shown on Figure 3.21. The coefficient estimates are listed in Table 3.6 (excluding the model based on $m_3 = 7.45$). Besides the model with $m_3 = 7.10$ (that is *Model4.nEMG*), the remaining three models are named *Model5.1.nMEG* through *Model5.3.nEMG*. Some discussions of this table and the resulting predictions are in order:

- A large fraction of the variability in c_0 estimates is attributed to the different m_3 to which estimate is anchored (vertical lines on Figure 3.21).
- Estimated coefficient c_1 (the semi-major axis length of the elliptical mean displacement profile) is not sensitive to the choice of m_3 . The model coefficients of σ and v parameters are also insensitive to the choice of m_3 . It makes sense that these m_3 -insensitive coefficients are not related to the \mathbf{M} -scaling part of the model.
- Coefficients of the σ_{eq} model are also not sensitive to m_3 , except for the case of $m_3 = 6.4$. It is possible that poorer fit to large earthquakes by the $m_3 = 6.4$ \mathbf{M} -scaling relation weakens the \mathbf{M} dependence of σ_{eq}^2 .
- The m_2 estimates are in the range of 0 (when $m_3 = 7.32$) to 2 (when $m_3 = 6.4$). This range is wide, yet conformed to the above-discussed constraints.
- Predicted quantiles coalesce but splinter outside data's \mathbf{M} range (Figure 3.22). The splintering is a result of having wide ranging m_1 and m_2 . The coalescence within data's \mathbf{M} range is helped by the limited magnitude range of data and the relatively gentle curvature of $f_M(\mathbf{M})$, with which the three \mathbf{M} -scaling coefficients (m_1, m_2, m_3) are able to trade off with each other to maintain a close fit to data.
- So far, we have not discussed contributions from the estimation uncertainty of m_1 and m_2 . Estimation uncertainty of m_1 is non-negligible, and its contribution to the prediction uncertainty at smaller magnitude ($\mathbf{M} < 6.5$) should be considered in PFDHA. The reported estimation uncertainty of m_2 is also non-negligible, but we judge that it does not need to be further considered because it is subject to the constraints discussed above.

- The four models with credible m_3 values cannot be distinguished by the test statistics provided by the GAMLSS modeling. This indicates the challenge in assigning statistic-based logic-tree weights for these alternative FDMs without adding more data outside the current data \mathbf{M} range, either empirical or simulated. Because of this, we do not attempt to assign statistics-based weighting to these models. If statistics-based weighting is desired, one option is to base the weighting scheme on the truncated normal distribution of m_3 after the left truncation point is adequately assessed.
- Logic-tree weights for these models may also be assigned by experts via a solicitation process, which is outside the scope of this study. We judge that, with limited basis, model with $m_3 = 6.4$ should be given the lowest weight among the four models (say, a weight of 0.1), and the remaining three models should be given equal weight (a weight of 0.3).

3.5.2 Uncertainty of other coefficients

We did not systematically assess the estimation uncertainty of model coefficients $c_1, c_{v1}, c_{v2}, c_{v3}, c_{v5},$ and c_{v6} . However, we postulate that these sources of uncertainty are probably unimportant compared to that associated with the uncertainty in \mathbf{M} -scaling.

3.6 DISCUSSIONS

3.6.1 Predicted Profile Shape of D : Vindication of Using Ellipse Equation for $\ln(D)$

The μ parameter of *Model4.nEMG* is assumed to follow an ellipse (the x^* term). This functional form, which is the theoretical profile of D of a simple elastic crack model, has often been used to model observed D , but not $\ln(D)$. P11 did not explain the rationale behind their choice of ellipse for use in one of their models for $\ln(D)$. Here, we try to provide some justifications by assessing its reasonableness against the profile shapes of D from other displacement data published in previous studies. On Figure 3.23, the predicted median of D by *Model4.nEMG* for an $\mathbf{M} = 7.5$ earthquake is plotted in linear y-scale against $l2L$. The plotted profile does not resemble an ellipse. In addition to the different curvature toward rupture tips, the predicted profile (black curve) has a narrower top than that of the elliptical profile. Our predicted profile is closely matched by the hyperbolic tangent function ($c_1 + c_2 * \tanh(c_3 * l2L)$) used in the recent FDM study by Youngs et al. (2021) for rail transportation system.

The predicted median by *Model4.nEMG* can also be reasonably approximated by a trapezoid (red dashed curve), although not as well as the match by the \tanh function described above. The good match by a trapezoid intrigues us because trapezoid is similar in shape to the (symmetric) slip profile of the small-scale yielding crack model (Kanninen and Popelar, 1985; Scholz, 2019, Section 1.1.4) where, roughly, displacements taper linearly toward the crack tips. In

addition, a near-linear slip taper is frequently observed in the slip profiles of earthquakes and faults (for examples, Scholz and Lawler, 2004; Manighetti et al., 2005). There are several theoretical explanations for the linear fault tip taper (see, for example, Martel and Shacat, 2006). The similarity of predicted shape of D to that of the small-scale yielding crack model and to those observed in past earthquakes justify our use of the ellipse equation for $\ln(D)$.

Unlike the crack models where theoretical displacement decreases to zero at the crack tips, our probability distribution of slip at the rupture tip spans a wide range of value, and at larger \mathbf{M} , only values at the very far left tail are small enough to be considered zero. As an example, at the rupture tip of an \mathbf{M} 7.5 event, the predicted 5th, 50th, and 95th percentiles are about 0.04 m, 0.5 m, and 2 m, respectively. Such a wide range of percentiles may have multiple causes, related to both data and model inadequacy. Data inadequacy may include incompletely mapped rupture trace, uncertainty in rupture tip location, under-reporting of small displacements in field surveys, exclusion of zero displacement in regression. As one example of model inadequacy, a different type of probabilistic distribution, such as a zero-inflated distribution or a more heavily left-skewed distribution, maybe more suitable for modeling the slip near the rupture tip.

3.6.2 Asymmetric slip profile

The parameters μ and ν of *Model4.nEMG* are symmetric about the midpoint of main trace (where $l2L = 0.5$), which, in turn, leads to symmetric profiles of predicted mean and quantiles of D . This assumed symmetry often contradicts the asymmetric profile observed in individual earthquakes or the composite of similar-shape profiles (for example, Manighetti et al., 2005). While recognizing the prevalence of asymmetric profile in the FDHI data, we adopt the symmetry assumption for its simplicity in model formulation and regression, as well as its ease of use in forward application. Given that each earthquake in the FDHI database has different profile shape and random binary skew direction, it is reasonable to assume that their composite, after accounting for the fixed-effect \mathbf{M} -scaling (f_M) and the earthquake-specific random intercept (δ_{eq_i}), can be adequately modeled by a symmetric function of $l2L$ (in our study, the equation of an ellipse).

To verify this symmetry assumption, the normalized quantile residuals from *Model4.nEMG* are plotted against $l2L$ in the top panel of Figure 3.24. Although these residuals do not show a visually detectable trend with $l2L$, the fitted line has a slightly negative slope of -0.134 with an estimation standard error of 0.06, yielding a p -value of 0.04. We attribute this weak linear trend to the fact that about 60% of our slip data come from right-skewed profiles. Out of curiosity, we replot the residual in the bottom panel of Figure 3.24, but this time we flip the orientation of left-skewed profiles so that profile of every earthquake is consistently skewed to the right. The replotted residuals show a clear linear downward trend.

Ignoring the slightly downward slope shown in the top panel of Figure 3.24, we conclude that our (spatially) symmetric model for μ is applicable when the skew direction of future slip profile is unknown *a priori* and treated as a binary random variate with equal probability. For cases

where profile's skew direction can be reliably forecasted (such as from geomorphic analysis of paleoslip histories), one may want to use a FDM with an asymmetric model for μ to take advantage of the benefits afforded by an asymmetric profile shape. For illustration, we derive such a model using slip profiles that are consistently right-skewed (data shown on the bottom panel of Figure 3.24). The formulation of the asymmetric μ is the same as Equation (3.14) except for the extra $(l2L_{ij} - 0.5)$ term:

$$\begin{aligned} \mu_{ij}(\mathbf{M}_i, l2L_{ij}) = & c_0 + \delta_{eq_i} + m_2(\mathbf{M}_i - 7.1) + \frac{(m_2 - m_1)}{c_n} \ln\left(\frac{1 + e^{-c_n(\mathbf{M}_i - 7.1)}}{2}\right) \\ & + c_1(x_{ij}^* - 1) + c_2(l2L_{ij} - 0.5) \end{aligned} \quad (3.16)$$

The linear $(l2L_{ij} - 0.5)$ term, prompted by the linear residual trend revealed in the bottom panel of Figure 3.24, may not be the most flexible way to model the asymmetry of average profile, but it is the simplest among the formulations that we are aware of. Estimated coefficients of this asymmetric FDM (*Model6.nEMG*) are tabulated in Table 3.1. The weaker $l2L_f$ -dependence in the ν parameter is consistent with the previously mentioned notion that imposing slip-profile symmetry on observed asymmetric profiles enhances the $l2L_f$ dependence.

The 50th and 95th percentiles predicted by the asymmetric model for an \mathbf{M} 7.5 rupture are shown on Figure 3.25. The predicted percentiles differ noticeably from those predicted by *Model4.nEMG* (the symmetric counterpart) for locations away from the center of main trace.

In addition to the differences in predicted quantiles, the asymmetric model differs from the symmetric model in two other aspects. First, the asymmetric slip profile peaks near $l2L = 0.4$. Secondly, although the fault-tip tapers are also roughly linear (the left panel of Figure 3.25), the left taper is now steeper than the right taper.

To fully develop and document the above asymmetric FDM, including asymmetric ν model, epistemic uncertainty, and implementation in PFDHA, further efforts would be needed but are outside the scope of this report.

3.6.3 Similarity between Different Stochastic Representations of Displacement

In the FDMs of Youngs et al. (2003) and Moss and Ross (2011), D is represented as the product of a log-normal variate (D_{ave} or D_{max}) and another random variate, which is either Weibull or gamma distributed D/D_{ave} or beta distributed D/D_{max} . For the purpose of demonstrating the similarity between different stochastic representations, we will ignore the fact that different displacement metrics maybe used for D .

A casual glance at the three representations of $\ln(D) = (G - E)$, $D = D_{ave} * \left(\frac{D}{D_{ave}}\right)$, and $D = D_{max} * \left(\frac{D}{D_{max}}\right)$, one may suggest that they are three very different

representations, but it is not so; they are actually similar when compared in the log-domain of D . For the convenience of subsequent discussions, we call them the 1st, the 2nd, and the 3rd representation, respectively. We also omit index i for earthquake and j for slip measurement in an earthquake.

The logarithmically transformed 2nd representation is

$$\ln(D) = \ln(D_{ave}) + \ln(D/D_{ave}) \quad (3.17)$$

In Section 3.6.3.2 below, it is shown that the normal variate G of the 1st representation is related to $\ln(D_{ave})$ of the 2nd representation in the following way,

$$G = \ln(D_{ave}) + G_{l2L,\ln D_{ave}} \quad (3.18)$$

and the 1st representation can be re-expressed as

$$\ln(D) = \ln(D_{ave}) + (G_{l2L,\ln D_{ave}} - E) \quad (3.19)$$

In Equations (3.17) and (3.19), $\ln(D)$ is written as the sum of the common variate $\ln(D_{ave})$ and a different second variate. The second variate is $\ln(D/D_{ave})$ in Equation (3.17) and $(G_{l2L,\ln D_{ave}} - E)$ in Equation (3.19). The types of probability distributions of the second variates are different: $\ln(D/D_{ave})$ is either a log-transformed Weibull variate or a log-transformed gamma variate according to Youngs et al. (2003) and Moss and Ross (2011), while $(G_{l2L,\ln D_{ave}} - E)$ is an nEMG variate. We will show later in Section 3.6.3.3 the probability distribution of $\ln(D/D_{ave})$ is matched well by the nEMG distribution, hence the 1st and the (log-transformed) 2nd representation are similar.

The same can be done for the 1st and the 3rd representations. After logarithmically transforming the 3rd representation, we have

$$\ln(D) = \ln(D_{max}) + \ln(D/D_{max}) \quad (3.20)$$

In Section 3.6.3.2 below, it is shown that component G of the 1st representation is related to $\ln(D_{max})$ of the 3rd representation in the follow way,

$$G = \ln(D_{max}) + G_{l2L,\ln D_{max}} \quad (3.21)$$

and the 1st representation can be re-expressed as

$$\ln(D) = \ln(D_{max}) + (G_{l2L, \ln D_{max}} - E) \quad (3.22)$$

In Equations (3.20) and (3.22), $\ln(D)$ is the sum of the common variate $\ln(D_{max})$ and a different second variate. The second variate is $\ln(D/D_{max})$ in Equation (3.20) and $(G_{l2L, \ln D_{max}} - E)$ in Equation (3.22). The types of probability distributions of the second variates differ: $\ln(D/D_{max})$ is a log-transformed beta variate according to Youngs et al. (2003) and Moss and Ross (20011), while $(G_{l2L, \ln D_{max}} - E)$ is an nEMG variate. We show later in Section 3.6.3.3 that the probability distribution of $\ln(D/D_{max})$ is matched well by the nEMG distribution. Hence, the 1st and the (log-transformed) 3rd representation are also similar.

3.6.3.1 Magnitude-scaling relations of $\ln(D_{ave})$ and $\ln(D_{max})$

The ability to re-express nEMG distribution's G component in terms of $\ln(D_{ave})$ or $\ln(D_{max})$ hinges on their similarity in \mathbf{M} -scaling relation. Using linear regression of $\ln(D_{ave})$ and $\ln(D_{max})$ data against the bilinear relation $(c_0 + f_M)$, we demonstrate that such similarity indeed exists. The D_{ave} data used herein are obtained using the envelop method (Haeussler, 2004), and D_{max} is simply the maximum of measured D at points along the main trace of an earthquake.

For $\ln(D_{ave})$ of the FDHI database, we obtain

$$\begin{aligned} \ln(D_{ave}) &\sim N(\mu_{\ln D_{ave}}, \sigma_{\ln D_{ave}}^2) \\ \mu_{\ln D_{ave}} &= 0.4761 + 0.8844 (\mathbf{M} - 7.0720) + \frac{0.8844 - 3.5964}{10} \ln(1 + e^{-10 * (\mathbf{M} - 7.0720)}) \\ \sigma_{\ln D_{ave}} &= 0.7064 \end{aligned} \quad (3.23)$$

For $\ln(D_{max})$ of the FDHI database, we obtain

$$\begin{aligned} \ln(D_{max}) &\sim N(\mu_{\ln D_{max}}, \sigma_{\ln D_{max}}^2) \\ \mu_{\ln D_{max}} &= 1.2662 + 0.8989 (\mathbf{M} - 7.0641) + \frac{0.8989 - 3.5151}{10} \ln(1 + e^{-10 * (\mathbf{M} - 7.0641)}) \\ \sigma_{\ln D_{max}} &= 0.7039 \end{aligned} \quad (3.24)$$

The fitted $\mu_{\ln D_{ave}}$, $\mu_{\ln D_{max}}$, and the $c_0 + f_M$ term of *Model4.nEMG* show similar **M**-scaling relations (Figure 3.26). Ignoring the slight difference in **M**-scaling rate, we may write

$$\mu_{\ln D_{ave}} \approx c_0 + f_M - 0.7642 \quad (3.25)$$

$$\mu_{\ln D_{max}} \approx c_0 + f_M + 0.0065 \quad (3.26)$$

which imply

$$\mu_{\ln D_{ave}} \approx \mu_{\ln D_{max}} - 0.7707 \quad (3.27)$$

The implied 1:1 relation between $\mu_{\ln D_{ave}}$ and $\mu_{\ln D_{max}}$ is corroborated by the $\ln(D_{ave}/D_{max})$ data shown on Figure 3.27, which exhibit a statistically insignificant upward trend with **M**. For reference, the average value of $\ln(D_{ave}/D_{max})$ is -0.8288, a tad more negative than the -0.7707 offset proposed in Equation (3.27). We have not found corroborating publications using different strike-slip dataset. Wells and Coppersmith (1994), who assumed a linear **M**-scaling relation, reported a scaling rate of 1.03 (± 0.08) and 0.90 (± 0.09) for $\log_{10}(D_{max})$ and $\log_{10}(D_{ave})$, respectively; the estimation standard error is reproduced inside the parenthesis. The $\ln(D_{ave}/D_{max})$ values from the Wells and Coppersmith strike-slip earthquakes are also plotted on Figure 3.27. An upward trend with **M** can be visually detected, consistent with their reported scaling rates. For reference, Wells and Coppersmith (1994) also reported, for all styles of faulting, a rate of 0.82 and 0.69 for $\log_{10}(D_{max})$ and $\log_{10}(D_{ave})$, respectively.

Based on the above regression results for $\ln(D_{max})$, one may postulate that the $(\delta_{eq,i} + c_0 + f_M(\mathbf{M}_i))$ term (which is equal to $\mu_{ij}(\mathbf{M}_i, l2L_{ij} = 0.5)$) can be a proxy of observed $\ln(D_{max})$, thus providing a much-needed physical interpretation for that term. To verify this postulate empirically, we plot the observed $\ln(D_{max})$ against the $(\delta_{eq,i} + c_0 + f_M(\mathbf{M}_i))$ term of *Model4.nEMG* on the left panel of Figure 3.28. They show a 1:1 relation with a 0.9863 correlation coefficient, and the scatter around the 1:1 line has a small standard deviation of 0.2548.

Although the plot supports our postulate, we are still uncomfortable with it because the exceedance probability of predicted $(\delta_{eq,i} + c_0 + f_M(\mathbf{M}_i))$ is about 0.1 (cumulative probability of about 0.9), based on *Model4.nEMG*. To further illustrate this concern, on Figure 3.29, we plot the probability density functions and several predicted percentiles along an **M** 7.2 rupture trace. The probability mass is about 0.1 between the predicted 0.999 quantile and the predicted μ . Are the observed $\ln(D_{max})$ data consistent with such a large exceedance probability? To shed light on this question, we use stochastic simulations to compute the probability of exceeding the realized maximum values taken from a given nEMG distribution. The simulations are carried out for $l2L = 0.5$ and for eight different sample sizes, ranging from 5 to 100 samples. To account for sampling variability, we repeat the calculations 100 times for each sample size. The results are shown on

Figure 3.30 as a function of sample size. With the sample size set to 100, the exceedance probabilities vary between 0.0001 and 0.0436, and their average is 0.0098, much smaller than the 0.1 probability noted above. With the sample size set to 10, exceedance probabilities vary considerably between 0.0005 and 0.3783 and the mean probability is 0.0939, close to 0.1. This begs the question whether the size of sampled slip data that lead to the observed D_{max} is around 10. Definition of such sample size is non-unique; for the sake of argument, we take it as the number of slip measurements within a 1-km window centered at the location of observed D_{max} . Using this definition, we find that the sample size of the majority of FDHI strike-slip earthquakes fall below 10 (subplot on Figure 3.3030), indicating the 0.1 probability of exceedance is possible.

Summarizing the above discussions, the interpretation of $(c_0 + f_M + \delta_{eq})$ as a proxy of observed $\ln(D_{max})$ holds empirically, but one need to be mindful of the 0.1 probability of exceeding the observed $\ln(D_{max})$. Given the finite exceedance probability, the adequacy of normalizing D to observed D_{max} (the 3rd representation) warrants further investigation. Instead of being the proxy of observed $\ln(D_{max})$, our preferred interpretation of $(c_0 + f_M + \delta_{eq})$ is that it corresponds roughly to the .90 quantile of $\ln(D)$ at a point.

Finally, we note that both $\sigma_{\ln D_{ave}}^2$ and $\sigma_{\ln D_{max}}^2$ are not constant; they decrease with increasing magnitude, similar to the trend of the variance σ_{eq}^2 of $\delta_{eq,i}$ discussed earlier. The \mathbf{M} -dependent functional form of Equation (3.5) is also applicable to $\sigma_{\ln D_{ave}}$ and $\sigma_{\ln D_{max}}$ and we have

$$\sigma_{\ln D_{ave}} = \max(1.1747 e^{-0.8860 * \max(\mathbf{M}-6.1, 0)}, 0.4) \quad (3.28)$$

$$\sigma_{\ln D_{max}} = \max(1.1391 e^{-0.8828 * \max(\mathbf{M}-6.1, 0)}, 0.4) \quad (3.29)$$

Such \mathbf{M} -dependent variances have not been reported in previous studies using a non-FDHI database.

3.6.3.2 Decomposing G Component

Based on the above results, we can now re-express G in terms of $\ln(D_{ave})$ or $\ln(D_{max})$. Remember that, from Equation (3.14), $G \sim N(c_0 + f_M + f_{l2L}, \sigma'^2 = \sigma^2 + \sigma_{eq}^2)$. Using the property $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, where $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are two independent normal variables, we re-write G as the sum

$$\begin{aligned} G &= \ln(D_{ave}) + G_{l2L, \ln D_{ave}} \\ G_{l2L, \ln D_{ave}} &\sim N(f_{l2L} + 0.7642, \sigma^2 + \sigma_{eq}^2 - \sigma_{\ln D_{ave}}^2) \end{aligned} \quad (3.30)$$

where $\ln(D_{ave})$ is independent of $G_{l2L, \ln D_{ave}}$. The 1st representation is now re-expressed as

$$\ln(D) = \ln(D_{ave}) + (G_{l2L, \ln D_{ave}} - E) \quad (3.31)$$

Similarly, we can re-write G as the sum

$$G = \ln(D_{max}) + G_{l2L, \ln D_{max}}$$

$$G_{l2L, \ln D_{max}} \sim N(f_{l2L} - 0.0065, \sigma^2 + \sigma_{eq}^2 - \sigma_{\ln D_{max}}^2) \quad (3.32)$$

and

$$\ln(D) = \ln(D_{max}) + (G_{l2L, \ln D_{max}} - E) \quad (3.33)$$

Note that $(G_{l2L, \ln D_{ave}} - E)$ and $(G_{l2L, \ln D_{max}} - E)$ are nEMG variates.

3.6.3.3 Approximation of $\ln(D/D_{ave})$ and $\ln(D/D_{max})$ by nEMG

Although the second random variate in the 2nd and the 3rd representations has a probability distribution different from the nEMG distribution, it can be matched reasonably well by the nEMG distribution in the logarithmic domain. This claim will be demonstrated below. This indicates that all three representations likely yield similar predictive distributions, if the same slip dataset and compatible model formulations of distribution parameters are used in regression analysis.

We sample 10,000 realizations of a Weibull variate whose distribution parameters are taken from the (D/D_{ave}) model of Moss and Ross (2011). We then fit an nEMG distribution to the logarithm of sampled values. This process is repeated for three different site locations: $l2L = 0, 0.35, \text{ and } 0.5$. The goodness of match by the nEMG distribution is examined using histograms and Q-Q plots (Figure 3.31). These plots show that the nEMG distribution matches the logarithm of Weibull variate reasonably well. Noticeable mismatches are mainly at the far tails.

The same analysis is carried out for the gamma-distributed (D/D_{ave}) , again using the distribution parameters from Moss and Ross (2011). The histograms and Q-Q plots are shown on Figure 3.32. The plots again indicate that the nEMG distribution matches the logarithm of gamma variate reasonably well, and noticeable mismatches are mainly at the far tails.

Incompatible support of probability distribution between the logarithm of beta-distributed D/D_{max} and $(G_{l2L, \ln D_{max}} - E)$ imposes a rigid constraint on $G_{l2L, \ln D_{max}}$ that warrants some discussions. The distribution of the logarithm of beta variate is supported on $(-\infty, 0]$,

whereas $(G_{l2L, \ln D_{max}} - E)$, an nEMG variate, is supported on $(-\infty, \infty)$. To match the logarithm of beta, $(G_{l2L, \ln D_{max}} - E)$ needs to be specialized to $-E$ by degenerating $G_{l2L, \ln D_{max}}$ to a constant value of 0. That is, as a random variable, $G_{l2L, \ln D_{max}}$ has zero mean and zero variance. The latter entails $\sigma^2 + \sigma_{eq}^2 - \sigma_{\ln D_{max}}^2 = 0$, but in practice a small non-zero value would also work, at the expense of a small probability mass for positive values.

Matched distributions at the three $l2L$ locations are shown on Figure 3.33, for which the fitted variance of $G_{l2L, \ln D_{max}}$ is non-zero but small. Large mismatch by the nEMG distribution (or, more precisely, by $-E$) again occurs primarily at the tails.

In a recent study, Moss et al. (2022) model the D/D_{max} of reverse earthquakes as a gamma variate, not as a beta variate. Using the distribution parameters of their D/D_{max} model, the logarithm of sampled D/D_{max} are matched well by the nEMG distribution. Because the logarithm of gamma-distributed (D/D_{max}) is supported on $(-\infty, \infty)$, degeneration of $G_{l2L, \ln D_{max}}$ is not needed. The diagnostic plots of the matched distribution are shown on Figure 3.34. The matched nEMG density function extends beyond $\ln(D/D_{max}) = 0$, but it can be truncated at 0 and then renormalized.

3.6.4 Probability Distribution of D

In this section, we investigate the types of probability distribution that can be utilized to characterize the D predicted by *Model4.nEMG* for $\ln(D)$. Our approach is similar to that used earlier for stochastic simulation. For a given rupture scenario, we take 10,000 random samples of $\ln(D)$ from the predicted nEMG distribution, then fit positively-supported probability distributions to the realized D values. We limit the candidate distributions to those available in the *gamlss.dist* package (Chapter 19, Rigby et al., 2020). Therefore, our search is comprehensive, but by no means exhaustive. The above process is repeated for several \mathbf{M} and $l2L$. The fitted distributions are then ranked by their Akaike information criteria (AIC).

The simulation results show that distributions GB2 (generalized beta type 2), BCPE (Box-Cox power exponential), BCT (Box-Cox t), and BCCG (Box-Cox Cole and Green) are often in the list of top-3 ranked distributions. These top-ranked distributions are either three-parameter or four-parameter distributions and have so far not been used for the modeling of D .

3.6.5 Range of Model Applicability

To understand how well our μ model extrapolates outside the magnitude range of the FDHI database, we turn to $\ln(D_{ave})$ and assess how well the $\ln(D_{ave})$ - \mathbf{M} scaling relation of Equation (3.23) fits three other D_{ave} datasets that include extra D_{ave} data from $\mathbf{M} > 7.9$ and $\mathbf{M} < 6.5$ earthquakes. This assessment is meaningful because $\ln(D_{ave})$ and $\ln(D)$ share the same \mathbf{M} -scaling relation of $c_0 + f_M$. The success of this assessment is hinged on the assumption that

similarity in \mathbf{M} -scaling between $\ln(D_{ave})$ of the new datasets and $c_0 + f_M$, if exists, extends outside the \mathbf{M} range of FDHI database. We collect D_{ave} data from the following three sources:

- Wells and Coppersmith (1994; WC94); D_{ave} and D_{max} of surface displacement from $5.55 \leq \mathbf{M} \leq 8.14$ earthquakes
- Wells and Youngs (2015; WY15); D_{ave} and \mathbf{M} digitized from a plot on their 2015 SSA poster; roughly, $5.51 \leq \mathbf{M} \leq 8.25$
- Anderson et al. (2021; AEL21); D_{ave} from $5.79 \leq \mathbf{M} \leq 8.35$ earthquakes

Data used in Takao et al. (2013) are unavailable to us.

All collected D_{ave} data are from strike-slip earthquakes except for WY15, which are from all styles of faulting (we do not have information to identify strike-slip earthquakes). The calculation method of D_{ave} used in each data source may be different. Hence, for an earthquake common to multiple datasets, D_{ave} values from different data sources may be different. The magnitude assigned to a common earthquake may also differ between data sources. It is not a good idea to merge these datasets into a single dataset.

The D_{ave} from FDHI and the above three data sources are plotted on Figure 3.35, one data source per panel. For reference, the median D_{ave} predicted by Equation (3.23), and the strike-slip model of WC94 are also plotted on Figure 3.35. In addition, we fit bilinear \mathbf{M} -scaling relation to the data from AEL21 and obtain

$$\ln(D_{ave}) = 0.5125 + 1.1662 (\mathbf{M} - 7.1361) + \frac{1.1662 - 1.5500}{10} \ln\left(\frac{1 + e^{-10(\mathbf{M}-7.1361)}}{2}\right) \quad (3.34)$$

Predicted median D_{ave} from Equation (3.34) is also plotted on Figure 3.35. The m_1 estimate (1.5500) from the AEL21 dataset is much smaller than that from the FDHI database, due to the large differences between the two datasets in the $\mathbf{M} < 7$ range. In contrast, the estimated m_2 (1.1662) is comparable to the m_2 estimates (Table 3.1) from the FDHI data. We do not fit bilinear relationship to the WY15 dataset because it contains earthquakes of all styles of faulting.

A novel feature of our FDM models (also Kuehn et al., 2022) is the bilinear \mathbf{M} scaling of displacement, in which m_1 is substantially larger than m_2 . This feature is not always clearly present in the above three non-FDHI datasets of D_{ave} . As shown by Equation (3.34), the AEL21 data hint at a bilinear scaling but the difference between their m_1 and m_2 is too small to provide a definite answer. Ignoring the fact that it contains earthquakes from every style of faulting, WY15 dataset also hints at a bilinear relationship (as shown by their loess smooth), but again the difference between the two scaling rates is too small to provide a definite answer. WC94 data do not require a bilinear scaling.

The four datasets examined here differ from each other, much more so in the $\mathbf{M} < 6.5$ range than in the $\mathbf{M} > 7$ range. Resolving the m_1 discrepancy is a challenge, given the state of available D_{ave} data. To reach a definite resolution, we will need additional earthquakes in the $\mathbf{M} < 6.5$ range and, more importantly, D_{ave} being computed by the same method (or different but compatible methods).

In spite of the inconsistency noted in moderate magnitudes, the $c_0 + f_M$ term of *Model4.nEMG* is able to fit every D_{ave} dataset reasonably well in the $\mathbf{M} > 7$ range. This is further illustrated by the data residuals (Figure 3.36). In addition, the $c_0 + f_M$ from *Model4.nEMG*'s epistemic variants (models corresponding to the credible range of m_3) are sufficiently wide to cover the m_2 implied by the four D_{ave} datasets, as shown on Figure 3.37.

Based on the above discussions, we assess that the applicable range of our FDM is $6.0 \leq \mathbf{M} \leq 8.3$. The 8.3 upper bound, which is 0.4 magnitude unit above the maximum \mathbf{M} of FDHI strike-slip earthquakes, is based on the ability of *Model4.nEMG* and its epistemic variants to fit D_{ave} data up to \mathbf{M} 8.3. Echoing the discussion in Section 3.5.1.2, when a moderate earthquake is the dominant seismic source in PFDHA, inclusion of estimation uncertainty in m_1 will be needed to reflect the lack of data. Broader epistemic uncertainty is also justified by the large observed discrepancy in D_{ave} noted above.

Table 3.1. Coefficients of Interim Updates and Preferred Model.

Model	Data Tier	Distribution	Regression	μ				σ_{eq}		σ		ν		τ		
				c_0	m_1	m_2	m_3	c_1	c_2	c_{v1}	c_{v2}	c_{v3}	c_{v4}	c_{v5}	c_{v6}	c_{v7}
Simple Update	1&2	Normal	lm	-19.0896	2.4211			1.9795		1.1831	-0.5050	1.1433				
<i>Model1.NO</i>	1	Normal	nlme	0.6235	3.6021	1.1081	7.0621	2.2248		1.1083	-0.6234	0.8812	-0.9515			
<i>Model2.SN</i>	1	Skew-Normal	gamlss	1.3808	3.5164	0.8122	7.1000	1.4123		1.0842	-0.7443	1.5127	-1.1603	-4.4219		
<i>Model3.ST</i>	1&2	Skew-t	gamlss	1.3511	3.4994	0.7947	7.1000	1.4026		1.1116	-0.7789	1.4096	-1.1459	-5.1167		4.1639
<i>Model4.nEMG</i>	1&2	nEMG	gamlss	1.3018	3.5070	0.7640	7.1000	1.3753		1.0681	-0.7548	0.2412		1.2234	-0.9565	
<i>Model6.nEMG</i>	1&2	nEMG	gamlss	1.3070	3.4761	0.8478	7.1000	1.6810	-0.6982	1.0468	-0.7418	0.2581		1.0051	-0.5072	

Notes:

$c_n = 10$;

Coefficient not used in a particular model is left blank;

Model4.nEMG is the preferred model;

Model6.nEMG is an extension of Model4.nEMG to allow an asymmetric profile for μ ; $\mu_{Model6} = \mu_{Model4} + c_2(l2L - 0.5)$; see Section 3.6.2 for details.

Table 3.2. Distribution Parameters and Percentiles Predicted by *Model1.NO*

M	Scenario			Distribution Parameters			Displacement Percentiles (m)				
	$l2L$	$l2L_f$	x^*	μ	σ_{eq}	σ'	5^{th}	16^{th}	50^{th}	84^{th}	95^{th}
5.8	0.05	0.05	0.43589	-5.00499	1.10833	1.41596	0.00065	0.00164	0.00670	0.02741	0.06884
7.0	0.05	0.05	0.43589	-0.78973	0.63240	1.08465	0.07624	0.15438	0.45397	1.33498	2.70302
8.0	0.05	0.05	0.43589	0.58056	0.40000	0.96775	0.36376	0.68261	1.78704	4.67837	8.77907
8.3	0.05	0.05	0.43589	0.91302	0.40000	0.96775	0.50723	0.95183	2.49184	6.52349	12.24148
5.8	0.50	0.50	1.00000	-3.74994	1.10833	1.41596	0.00229	0.00575	0.02352	0.09615	0.24150
7.0	0.50	0.50	1.00000	0.46532	0.63240	1.08465	0.26746	0.54155	1.59253	4.68312	9.48222
8.0	0.50	0.50	1.00000	1.83561	0.40000	0.96775	1.27609	2.39461	6.26895	16.41175	30.79705
8.3	0.50	0.50	1.00000	2.16807	0.40000	0.96775	1.77937	3.33904	8.74139	22.88445	42.94323

Note:

σ' is the distribution parameter after marginalizing the conditional normal distribution over $\delta_{eq,i}$.

Table 3.3. Distribution Parameters and Percentiles Predicted by *Model2.SN*

M	Scenario			Distribution Parameters				Displacement Percentiles (m)				
	$l2L$	$l2L_f$	x^*	μ	σ_{eq}	σ'	ν'	5 th	16 th	50 th	84 th	95 th
5.8	0.05	0.05	0.43589	-3.79977	1.08419	1.79247	-1.23319	0.00067	0.00183	0.00776	0.02954	0.06806
7.0	0.05	0.05	0.43589	0.33522	0.55487	1.53146	-2.18230	0.06950	0.16261	0.50800	1.28418	2.18625
8.0	0.05	0.05	0.43589	1.50253	0.40000	1.48239	-2.73499	0.24588	0.55971	1.66387	3.82376	6.01028
8.3	0.05	0.05	0.43589	1.74623	0.40000	1.48239	-2.73499	0.31374	0.71417	2.12302	4.87896	7.66886
5.8	0.50	0.50	1.00000	-3.00305	1.08419	1.37570	-0.75075	0.00344	0.00774	0.02605	0.08515	0.18259
7.0	0.50	0.50	1.00000	1.13194	0.55487	1.01240	-1.41077	0.42660	0.75136	1.66415	3.41923	5.32625
8.0	0.50	0.50	1.00000	2.29925	0.40000	0.93653	-1.87094	1.58995	2.67487	5.42265	9.87971	14.08352
8.3	0.50	0.50	1.00000	2.54294	0.40000	0.93653	-1.87094	2.02871	3.41302	6.91906	12.60609	17.96995

Note:

σ' and ν' are the distribution parameters after marginalizing the conditional SN distribution over $\delta_{eq,i}$.

Table 3.4. Distribution Parameters and Percentiles Predicted by *Model3.ST*

M	Scenario			Distribution Parameters					Displacement Percentiles (m)				
	$l2L$	$l2L_f$	x^*	μ	σ_{eq}	σ	ν	τ	5 th	16 th	50 th	84 th	95 th
5.8	0.05	0.05	0.43589	-3.80189	1.11160	1.33110	-5.11667	4.16393	0.00059	0.00230	0.00837	0.01728	0.02332
7.0	0.05	0.05	0.43589	0.31268	0.55146	1.33110	-5.11667	4.16393	0.03593	0.14077	0.51268	1.05815	1.42798
8.0	0.05	0.05	0.43589	1.46255	0.40000	1.33110	-5.11667	4.16393	0.11345	0.44450	1.61892	3.34138	4.50923
8.3	0.05	0.05	0.43589	1.70099	0.40000	1.33110	-5.11667	4.16393	0.14400	0.56419	2.05484	4.24111	5.72342
5.8	0.50	0.50	1.00000	-3.01068	1.11160	0.79483	-5.11667	4.16393	0.00561	0.01267	0.02742	0.04227	0.05056
7.0	0.50	0.50	1.00000	1.10389	0.55146	0.79483	-5.11667	4.16393	0.34334	0.77602	1.67911	2.58821	3.09538
8.0	0.50	0.50	1.00000	2.25375	0.40000	0.79483	-5.11667	4.16393	1.08417	2.45049	5.30223	8.17296	9.77449
8.3	0.50	0.50	1.00000	2.49220	0.40000	0.79483	-5.11667	4.16393	1.37611	3.11033	6.72995	10.37368	12.40645

Notes:

μ , σ , and ν are the distribution parameters of the conditional ST distribution, not the marginalized distribution;
 Displacement percentiles are computed using stochastic simulations.

Table 3.5. Distribution Parameters and Percentiles Predicted by *Model4.nEMG*

M	Scenario			Distribution Parameters				Displacement Percentiles (m)				
	$l2L$	$l2L_f$	x^*	μ	σ_{eq}	σ'	ν	5 th	16 th	50 th	84 th	95 th
5.8	0.05	0.05	0.43589	-3.84298	1.06811	1.09500	1.16626	0.00039	0.00155	0.00779	0.02981	0.06733
7.0	0.05	0.05	0.43589	0.27948	0.54148	0.59277	1.16626	0.03456	0.13421	0.52608	1.28890	2.10759
8.0	0.05	0.05	0.43589	1.40364	0.40000	0.46709	1.16626	0.11262	0.43729	1.67109	3.69344	5.56353
8.3	0.05	0.05	0.43589	1.63286	0.40000	0.46709	1.16626	0.14164	0.54995	2.10160	4.64495	6.99682
5.8	0.50	0.50	1.00000	-3.06713	1.06811	1.09500	0.75834	0.00224	0.00610	0.02323	0.07931	0.17192
7.0	0.50	0.50	1.00000	1.05532	0.54148	0.59277	0.75834	0.23502	0.57037	1.51373	3.25210	5.11827
8.0	0.50	0.50	1.00000	2.17949	0.40000	0.46709	0.75834	0.78966	1.90862	4.76817	9.16106	13.29835
8.3	0.50	0.50	1.00000	2.40871	0.40000	0.46709	0.75834	0.99310	2.40033	5.99656	11.52116	16.72431

Note:

σ' is the distribution parameters after marginalizing the conditional SN distribution over $\delta_{eq,i}$.

Table 3.6. Coefficients of nEMG Distributional Models, Conditional on Different Credible m_3

	μ				σ_{eq}		σ	ν		
	c_0	m_1	m_2	m_3	c_1	c_{v1}	c_{v2}	c_{v3}	c_{v5}	c_{v6}
<i>Model5.1.nEMG</i>	1.78647	3.23756	0.00000	7.32	1.37752	1.08283	-0.72562	0.24192	1.22189	-0.95510
<i>Model4.nEMG</i>	1.30182	3.50698	0.76397	7.10	1.37534	1.06811	-0.75483	0.24120	1.22339	-0.95650
<i>Model5.2.nEMG</i>	0.42841	4.08112	1.51908	6.75	1.38066	1.08784	-0.71501	0.24258	1.22104	-0.95441
<i>Model5.3.nEMG</i>	-0.59189	5.31589	1.95791	6.40	1.37705	1.06887	-0.59671	0.24148	1.22291	-0.95638

Note:

m_3 is fixed in regression analysis.

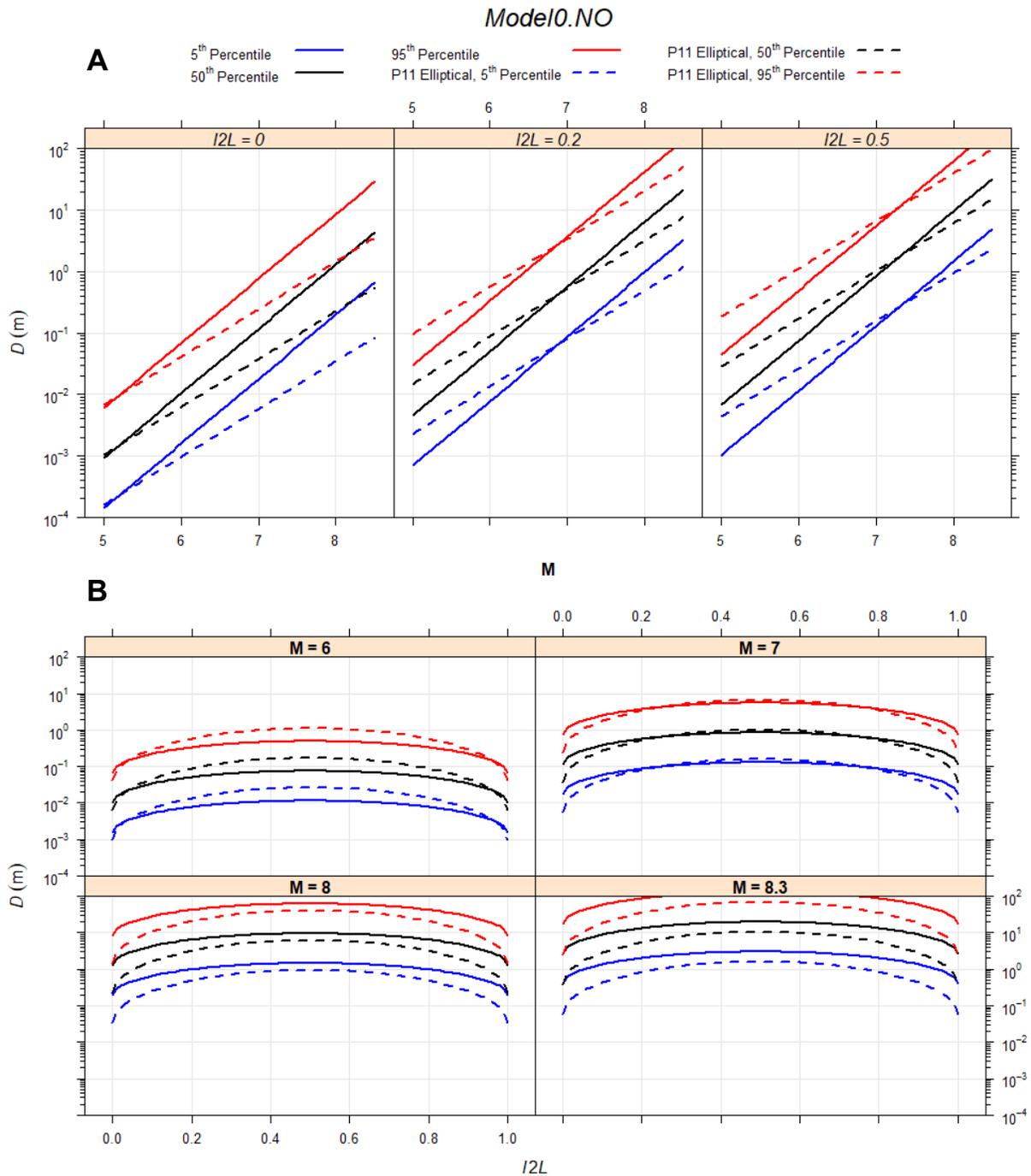


Figure 3.1. Quantiles of fault displacement (D) predicted by the simple update. (A) Displacement plotted versus magnitude (M). For comparison, quantiles predicted by the original Petersen et al. (2011, P11) elliptical model are plotted as dashed lines. (B) Displacement plotted against the normalized position along the main trace ($l2L$). For comparison, quantiles predicted by P11 (Petersen et al., 2011) are plotted as dashed curves.

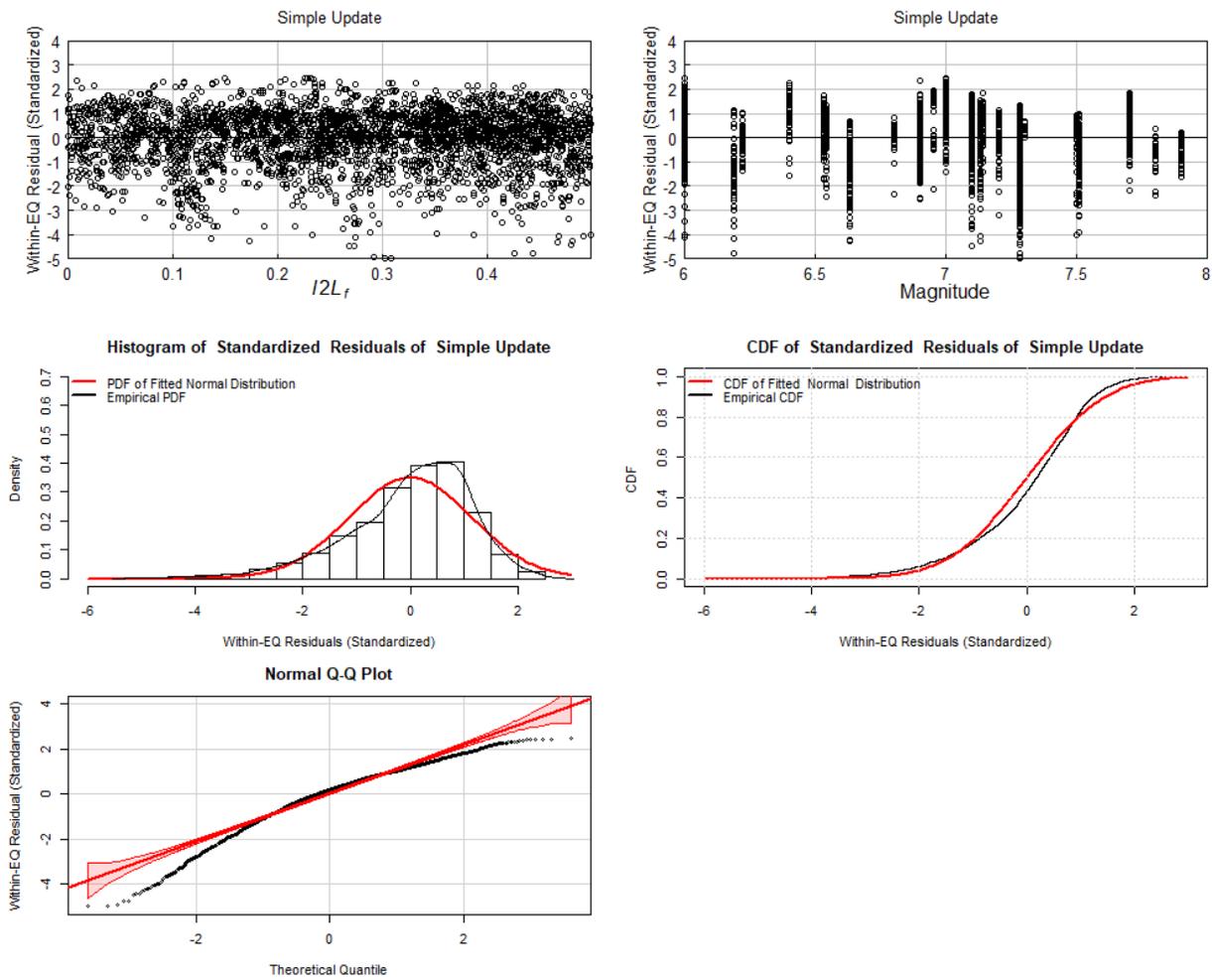


Figure 3.2. Residual diagnostic plots of the simple update to Petersen et al. (2011). Standardized residual ϵ/σ , where $\epsilon = \text{observed } \ln(D) - \text{predicted mean of } \ln(D)$ and $\sigma = \text{standard deviation of } \epsilon$, is used on this figure. (Top left) residuals versus $l2L_f$. Variable $l2L_f$ is the $l2L (= l/L)$ folded at the midpoint ($l/L = 0.5$) of the main trace, l is the along-main-trace position of the data point, and L is the length of the main trace. (Top right) residuals versus magnitude (M). (Middle left) histogram of residuals; probability density function (PDF) of the normal distribution fitted to the residuals is plotted as the red curve. (Middle right) empirical cumulative distribution function (CDF) of residuals; CDF of the normal distribution fitted to the residuals is plotted as the red curve. (Bottom left) quantile-quantile (Q-Q) plot that compares the quantiles of standardized residuals against the quantiles of standard normal distribution; the solid red line passes through the 1st and the 3rd quartiles, and the shaded area denotes roughly the 0.95 confidence level if residuals are taken from the standard normal distribution.

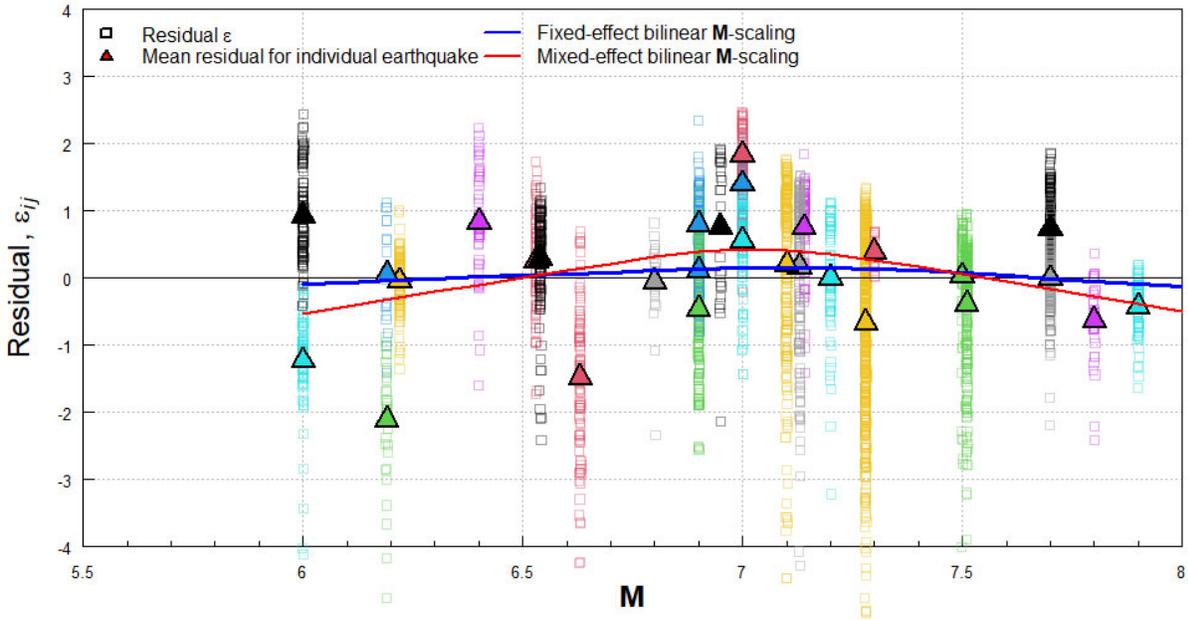


Figure 3.3. Data residuals of the simple update to Petersen et al., (2011), plotted versus magnitude (M). A residual is color coded according to the earthquake from which it is sampled. Solid triangle marks the mean of residuals in a particular earthquake ($\bar{\varepsilon}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \varepsilon_{ij}$, where n_i is the number of displacement measurements in earthquake i). The blue curve shows the result of a fixed-effect regression of the bilinear M -scaling function (function f_M of Equation (3.2)) to all data residuals. The red curve shows the result of a random-intercept mixed-effect regression, in which variation in the mean of individual earthquakes is modeled. The striking difference between these two fitted curves highlights the importance of proper weighting of individual earthquake afforded by the mixed-effect regression.

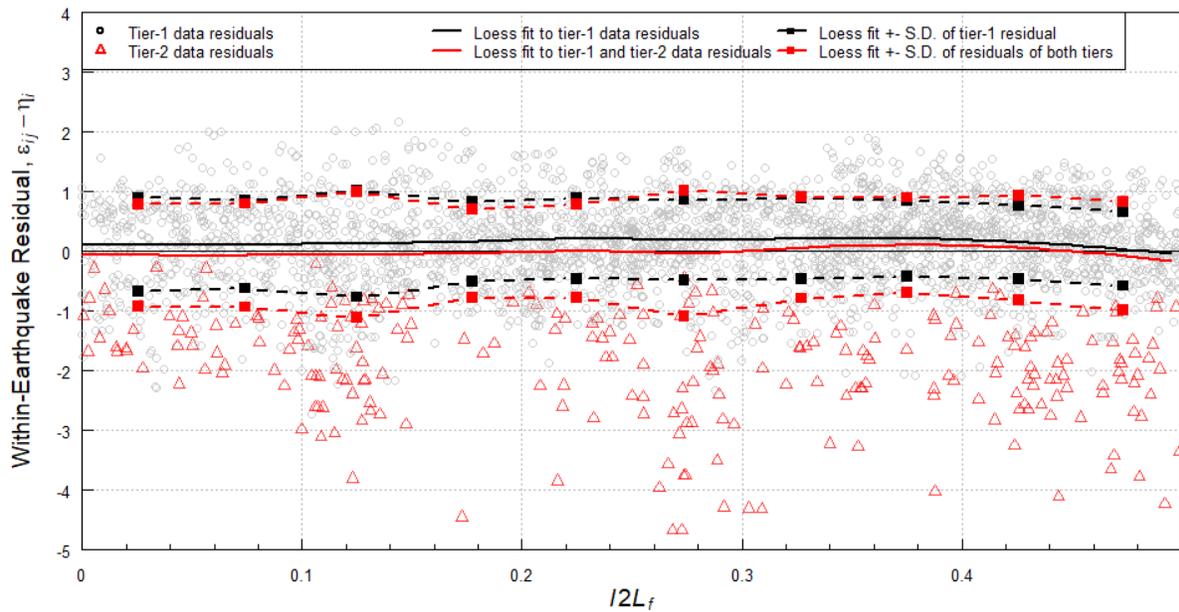


Figure 3.4. Within-earthquake residuals of the simple update to P11 (Petersen et al., 2011), plotted against $l2L_f$. Variable $l2L_f$ is the $l2L$ ($= l/L$) folded at the midpoint ($l/L = 0.5$) of the main trace, l is the along-main-trace position of the data point, and L is the length of the main trace. Within-earthquake residual is approximately computed as $(\varepsilon_{ij} - \bar{\varepsilon}_i)$ where $\bar{\varepsilon}_i$ is the mean of ε_{ij} in earthquake i , as defined in the text and in the caption of Figure 3.3. To visualize the variation of mean residual along the main trace, a loess smooth to all of the within-earthquake residuals is plotted as the red curve. A separate loess smooth to the tier-1 data residual is shown as the black curve. A comparison of these two loess smooths indicates that, relative to the tier-1 smooth, the estimated mean is decreased by an average of about 17%, due to the inclusion of tier-2 data. To assess its variation along the main trace, residual standard deviation (S.D.) in ten $l2L_f$ bins are computed and tag onto the loess smooths. Relative to tier 1's standard deviation, the standard deviation of residuals in an $l2L_f$ bin is increased by an average of about 27% along the main trace, due to the inclusion of tier-2 data. Tier-1 data's residual standard deviation reveals a slight downward trend with $l2L_f$. An assessment of the $l2L_f$ -dependence of residual standard deviation for tier-1 data is shown in Figure 3.5.

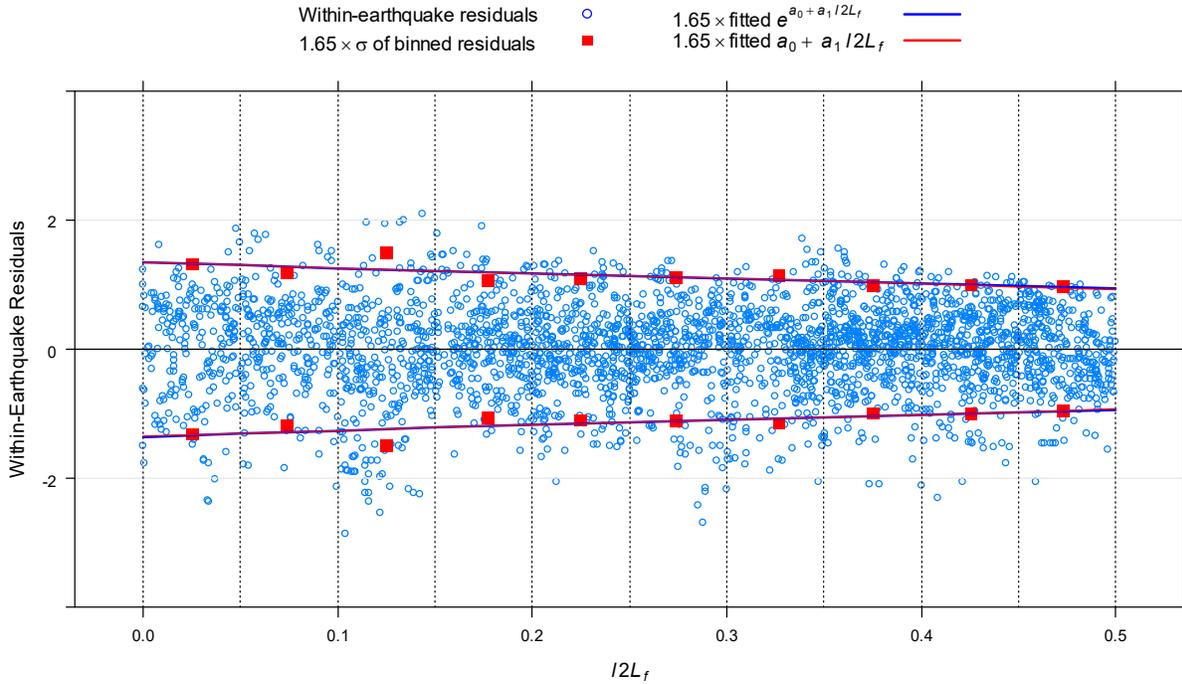


Figure 3.5. Within-earthquake residuals of *Model1.NO*, plotted against $l2L_f$. Variable $l2L_f$ is the $l2L (= l/L)$ folded at the midpoint ($l/L = 0.5$) of the main trace, l is the along-main-trace position of the data point, and L is the length of the main trace. Solid red symbol marks the standard deviation of residuals in a particular $l2L_f$ bin multiplied by 1.65. Bin boundaries are shown as the vertical dotted lines. Exponential function $e^{a_0 + a_1 l2L_f}$ and linear functions $a_0 + a_1 l2L_f$ are fitted to the bin standard deviations. The fitted exponential and linear functions, multiplied by 1.65, are shown as the blue and the red curves, respectively.

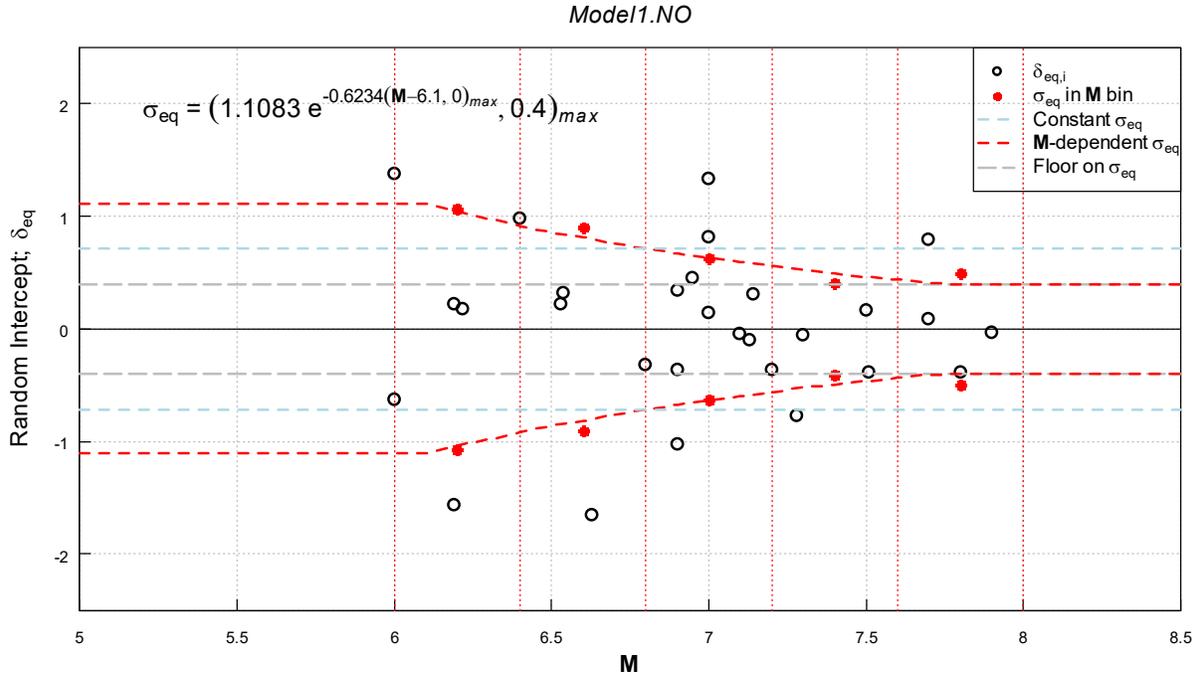


Figure 3.6. Random intercept $\delta_{eq,i}$ of *Model1.NO*, plotted versus magnitude (**M**). Solid red circles mark the standard deviation of $\delta_{eq,i}$ in respective magnitude bins bounded by the vertical dotted red lines. The horizontal short-dashed line marks the estimated σ_{eq} assuming it is a constant. The fitted **M**-dependent σ_{eq} model (Equation (3.4)) is written in the top left corner of the figure and shown as the red dashed curve. The horizontal long-dashed line indicates the imposed floor level of σ_{eq} in the large magnitude range.

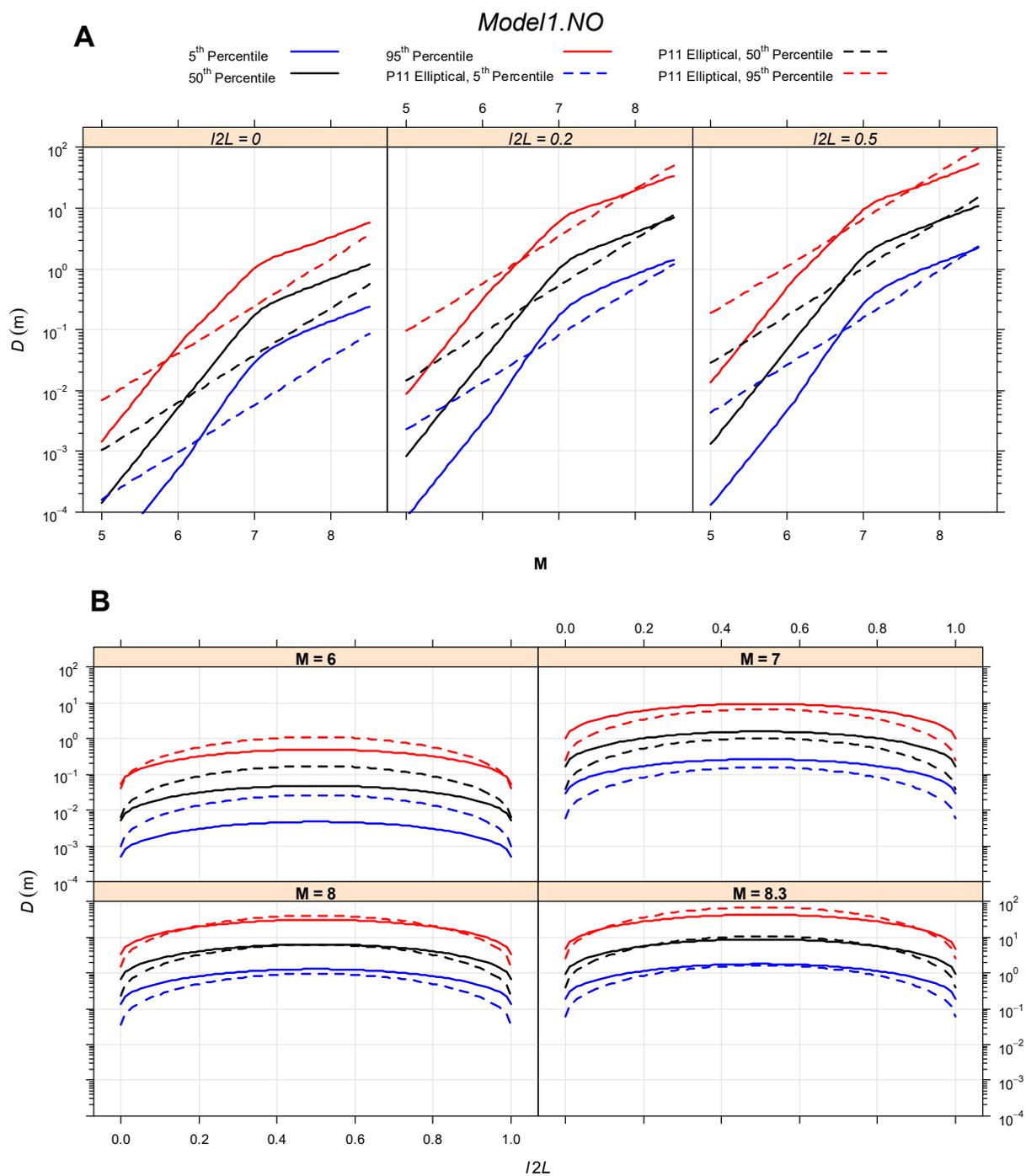


Figure 3.7. The 0.05, 0.50, and 0.95 quantiles (the 5th, 50th, and 95th percentiles in the figure title) of fault displacement (D) from the compound distribution of *Model1.NO*. (A) Quantiles versus magnitude M . (B) Quantiles versus the normalized position along the main ($l2L$). Quantiles predicted by P11 (Petersen et al., 2011) are shown as the long-dashed curves.

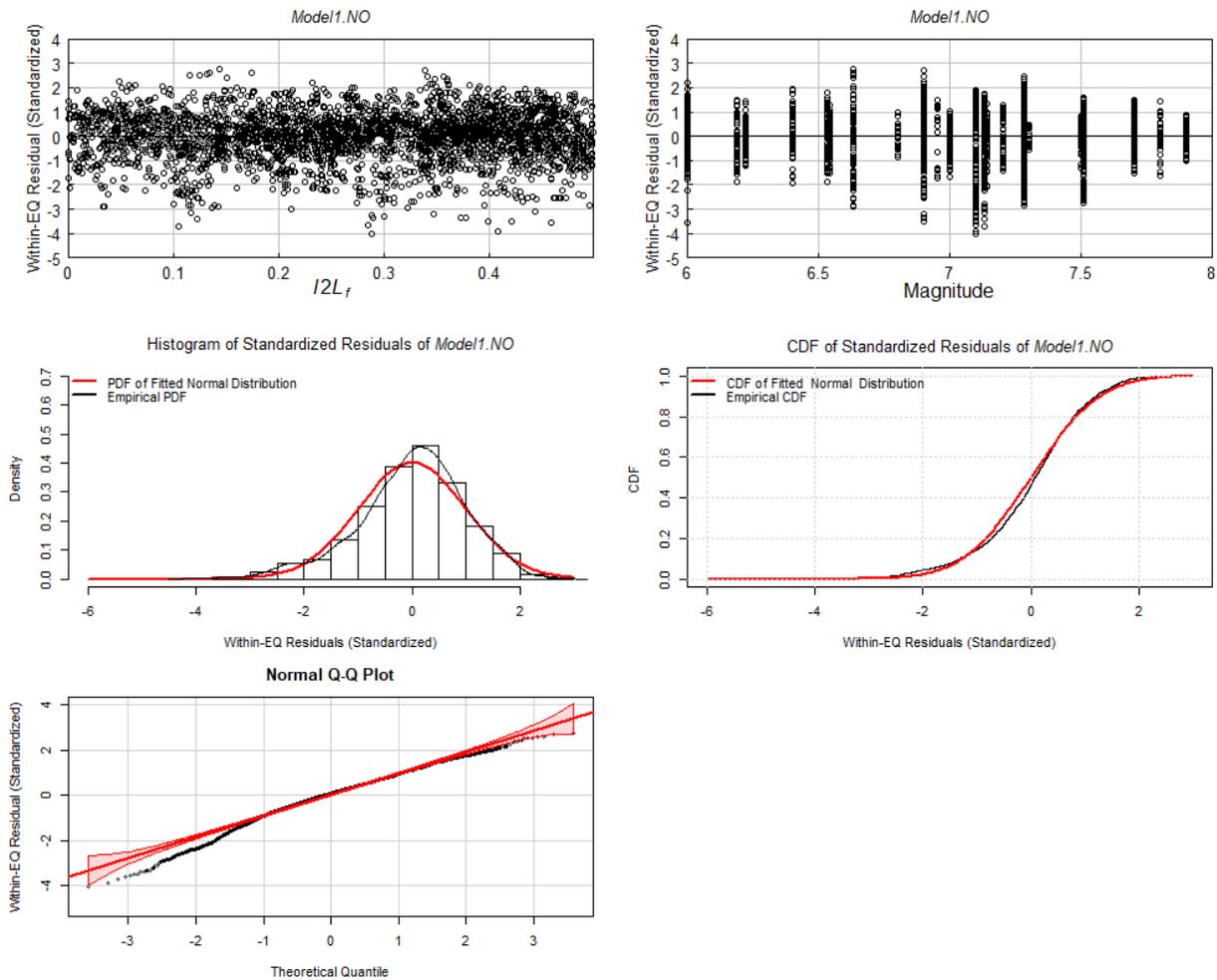


Figure 3.8. Residual diagnostic plots of *Model1.NO*. Within-earthquake (EQ) standardized residuals are used on this figure. (Top left) residuals versus $l2L_f$. Variable $l2L_f$ is the $l2L$ ($= l/L$) folded at the midpoint ($l/L = 0.5$) of the main trace, l is the along-main-trace position of the data point, and L is the length of the main trace. (Top right) residuals versus magnitude (**M**). (Middle left) histogram of residuals; probability density function (PDF) of the normal distribution fitted to the residuals is plotted as the red curve. (Middle right) empirical cumulative distribution function (CDF) of residuals; CDF of the normal distribution fitted to the residuals is plotted as the red curve. (Bottom left) quantile-quantile (Q-Q) plot that compares the quantiles of standardized residuals against the quantiles of standard normal distribution; the solid red line passes through the 1st and the 3rd quartiles, and the shaded area denotes roughly the 0.95 confidence level if residuals are taken from the standard normal distribution.

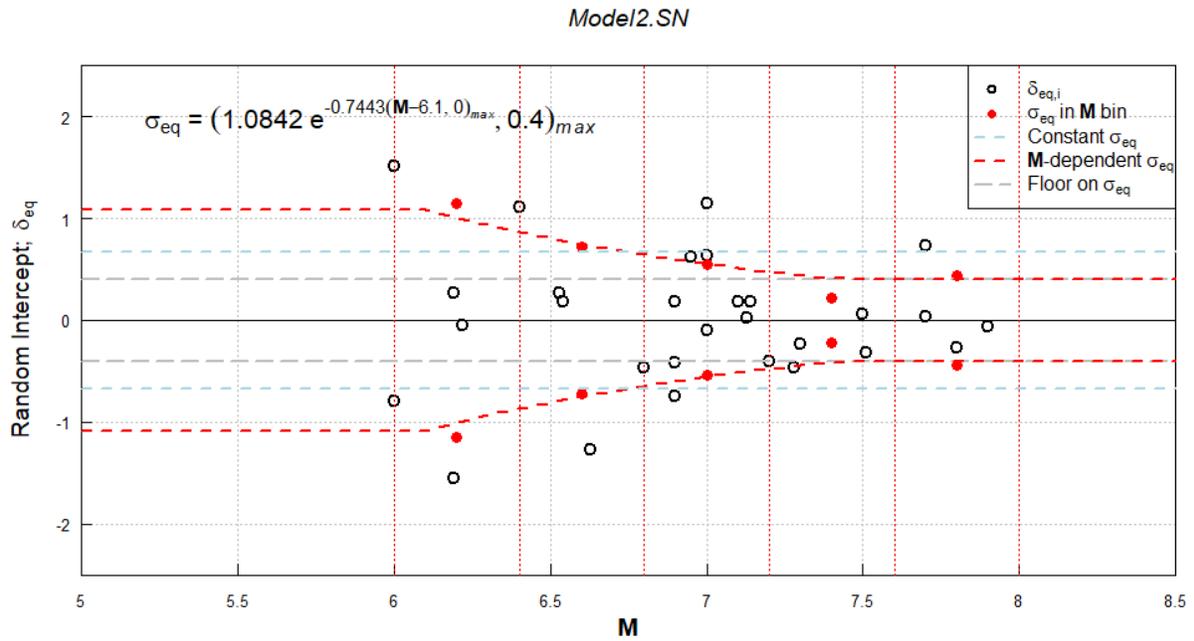


Figure 3.9. Estimated random intercept $\delta_{eq,i}$ of *Model2.SN*, plotted versus magnitude (**M**). Solid red circles are the standard deviation of $\delta_{eq,i}$ in respective magnitude bins bounded by the vertical dotted red lines. The short-dashed line marks the estimated σ_{eq} assuming it is a constant. The fitted **M**-dependent σ_{eq} model is shown as the red dashed curve; the model is also written in the top left of this figure. The long-dashed line indicates the imposed floor level of σ_{eq} .

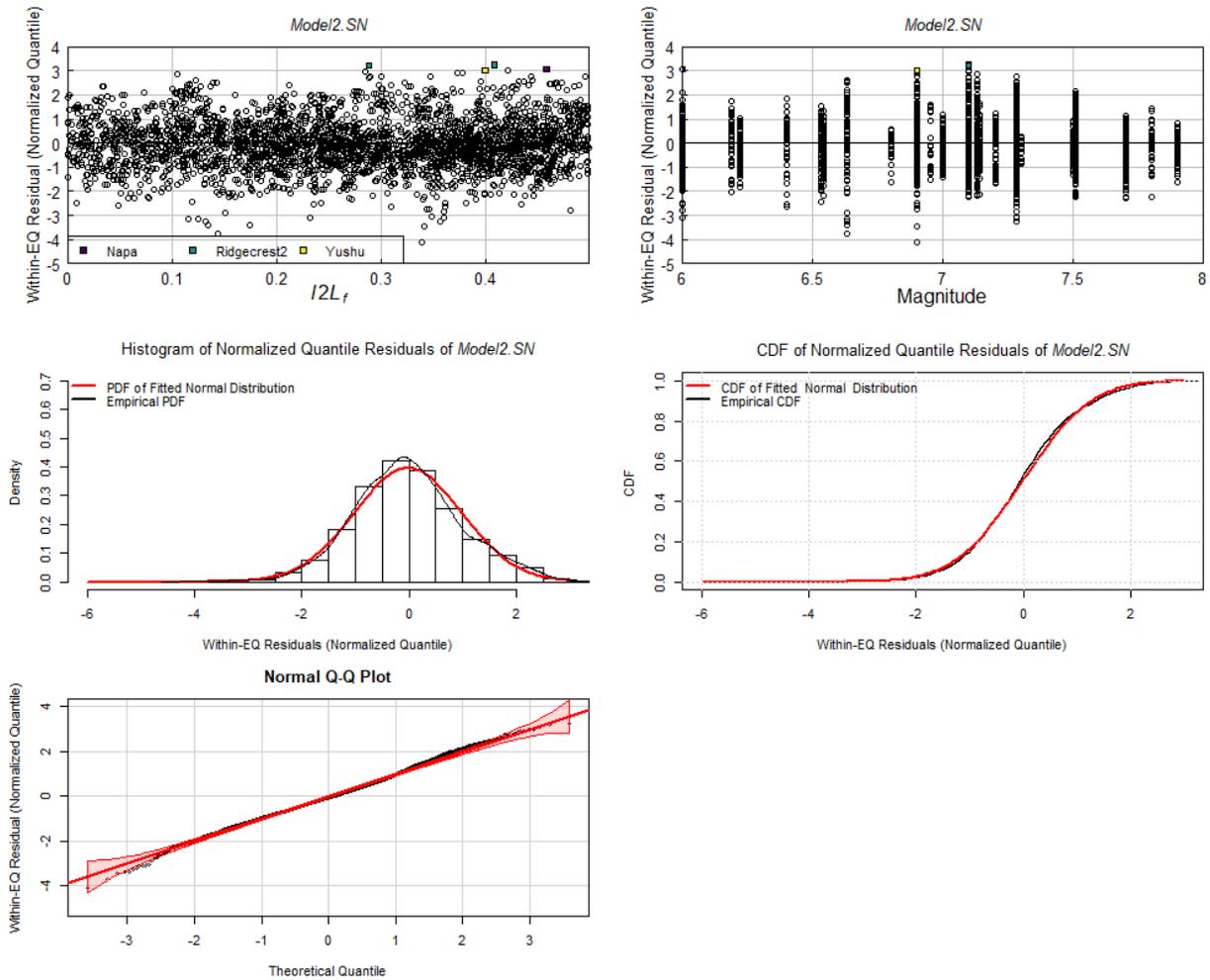


Figure 3.10. Residual diagnostic plots of *Model2.SN*. Within-earthquake (EQ) normalized quantile residuals are used on this figure. (Top left) residuals versus $l2L_f$. Variable $l2L_f$ is the $l2L$ ($= l/L$) folded at the midpoint ($l/L = 0.5$) of the main trace, l is the along-main-trace position of the data point, and L is the length of the main trace. (Top right) residuals versus magnitude (M). (Middle left) histogram of residuals; probability density function (PDF) of the normal distribution fitted to the residuals is plotted as the red curve. (Middle right) empirical cumulative distribution function (CDF) of residuals; CDF of the normal distribution fitted to the residuals is plotted as the red curve. (Bottom left) quantile-quantile (Q-Q) plot that compares the quantiles of normalized quantile residuals against the quantiles of standard normal distribution; the solid red line passes through the 1st and the 3rd quartiles, and the shaded area denotes roughly the 0.95 confidence level if residuals are taken from the standard normal distribution.

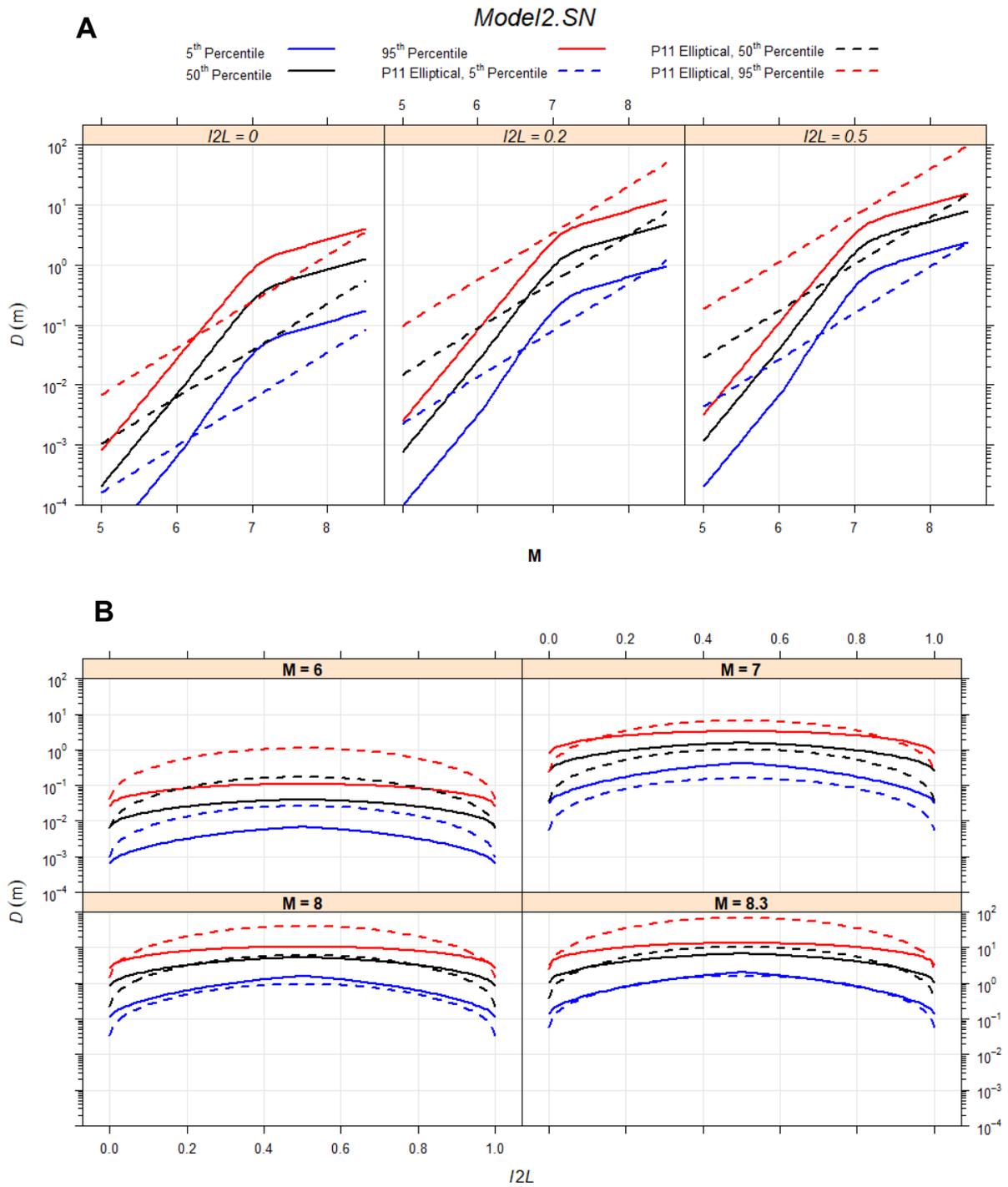


Figure 3.11. The 0.05, 0.50, and 0.95 quantiles (the 5th, 50th, and 95th percentiles in the figure title) of fault displacement (D) from the compound distribution of *Model2.SN*. (A) Quantiles versus magnitude M . (B) Quantiles versus the normalized position along the main trace ($l2L$). Quantiles predicted by P11 (Petersen et al., 2011) are shown as the long-dashed curves.

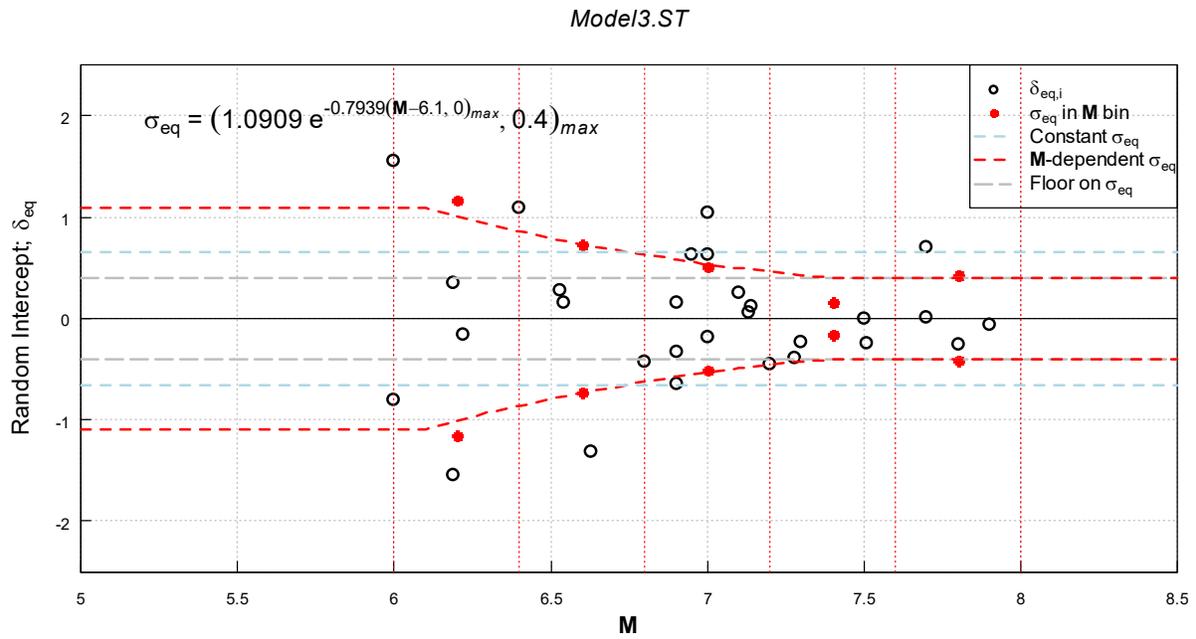


Figure 3.12. Estimated random intercept $\delta_{eq,i}$ of *Model3.ST*, plotted versus magnitude (**M**). Solid red circles are the standard deviations of $\delta_{eq,i}$ in respective magnitude bins bounded by the vertical dotted red lines. The short-dashed line marks the estimated σ_{eq} assuming it is a constant. The fitted **M**-dependent σ_{eq} model is shown as the red dashed curve; the model is also written in the top left of this figure. The long-dashed line indicates the imposed floor level of σ_{eq} .

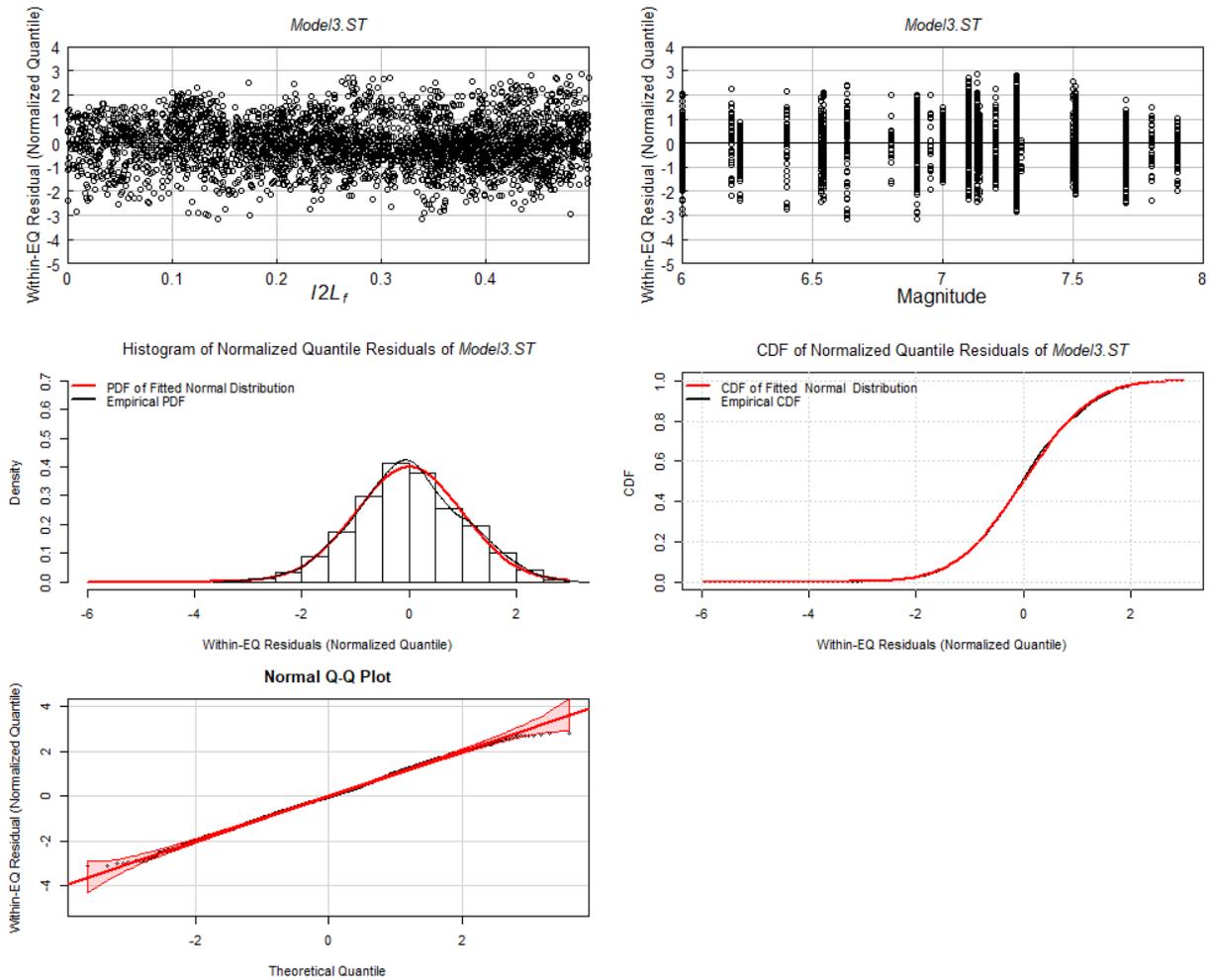


Figure 3.13. Residual diagnostic plots of *Model3.ST*. Within-earthquake (EQ) normalized quantile residuals are used on this figure. (Top left) residuals versus $l2L_f$. Variable $l2L_f$ is the $l2L (= l/L)$ folded at the midpoint ($l/L = 0.5$) of the main trace, l is the along-main-trace position of the data point, and L is the length of the main trace. (Top right) residuals versus magnitude (M). (Middle left) histogram of residuals; probability density function (PDF) of the normal distribution fitted to the residuals is plotted as the red curve. (Middle right) empirical cumulative distribution function (CDF) of residuals; CDF of the normal distribution fitted to the residuals is plotted as the red curve. (Bottom left) quantile-quantile (Q-Q) plot that compares the quantiles of normalized quantile residuals against the quantiles of standard normal distribution; the solid red line passes through the 1st and the 3rd quartiles, and the shaded area denotes roughly the 0.95 confidence level if residuals are taken from the standard normal distribution.

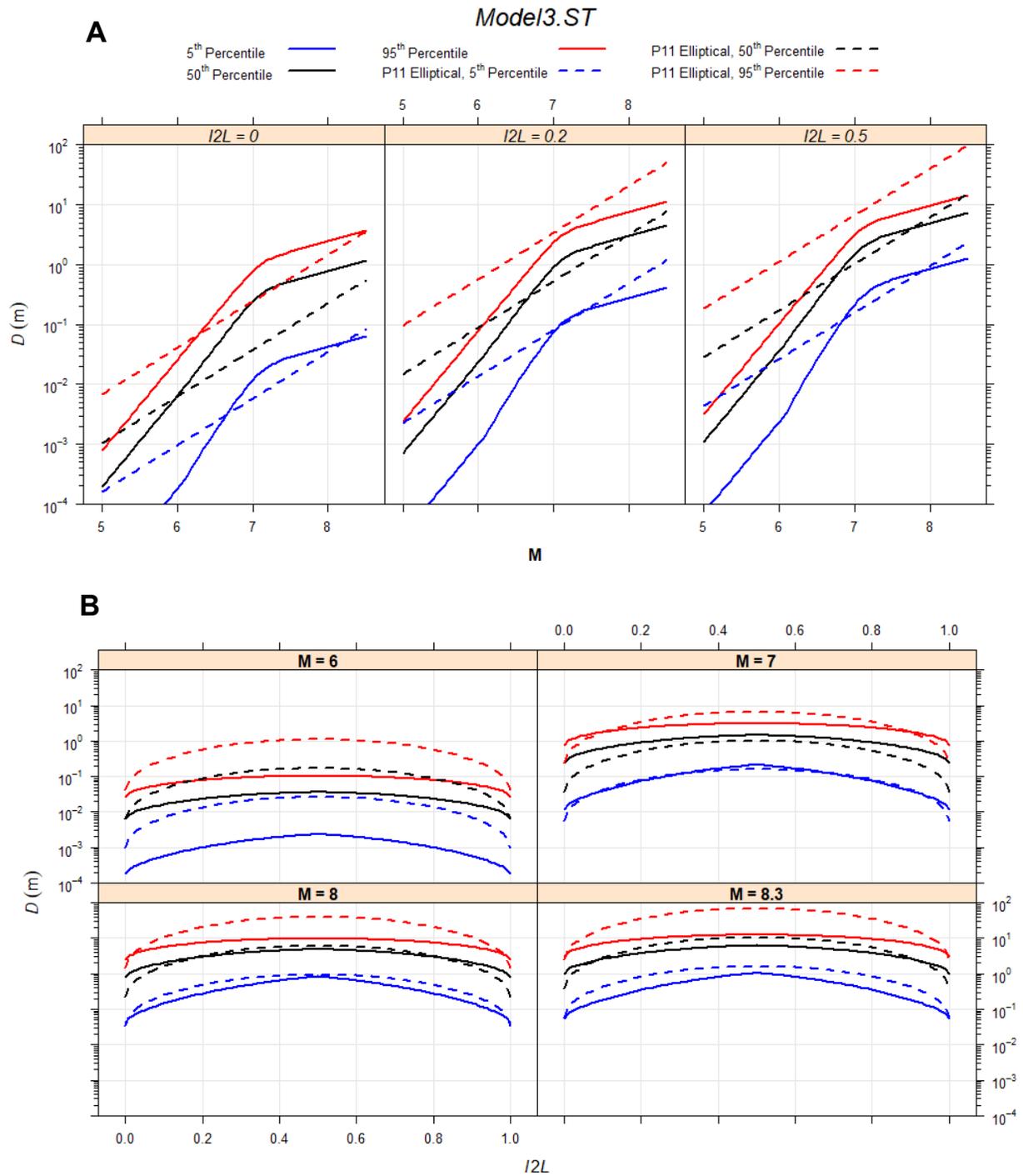


Figure 3.14. The 0.05, 0.50, and 0.95 quantiles (the 5th, 50th, and 95th percentiles in the figure title) of fault displacement (D) from the compound distribution of *Model3.ST*. (A) Quantiles versus magnitude M . (B) Quantiles versus the normalized position along the main trace ($l2L$). Quantiles predicted by P11 (Petersen et al., 2011) are shown as the long-dashed curves.

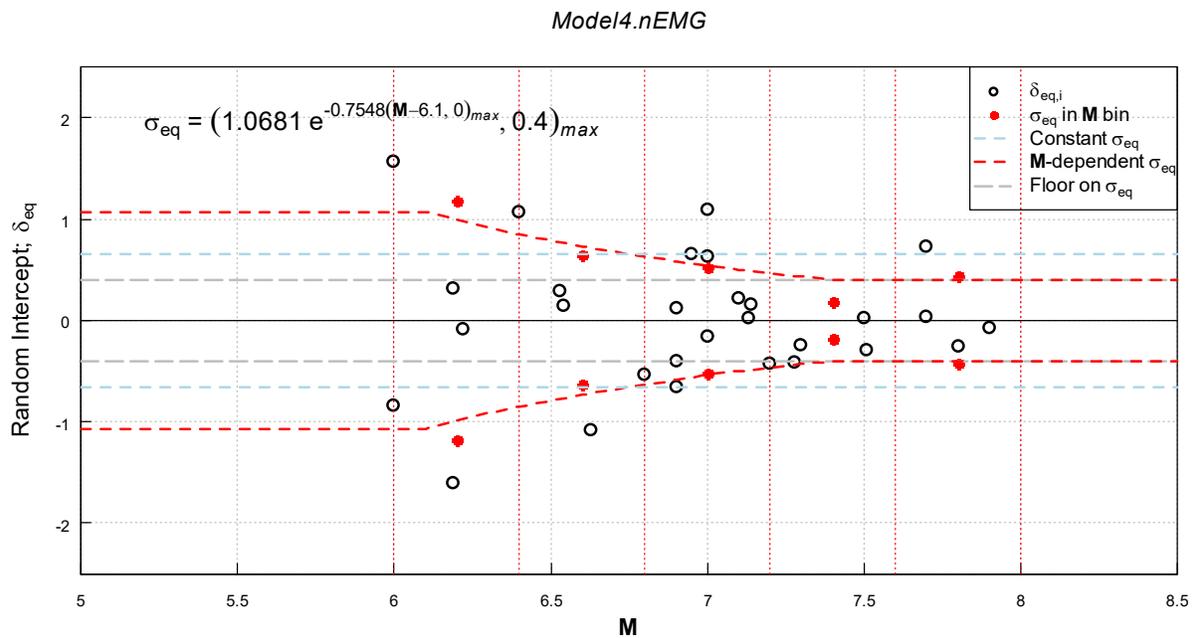


Figure 3.15. Estimated random intercept $\delta_{eq,i}$ of *Model4.nEMG*, plotted versus magnitude (**M**). Solid red circles are the standard deviations of $\delta_{eq,i}$ in respective magnitude bins bounded by the vertical dotted red lines. The short-dashed line marks the estimated σ_{eq} assuming it is a constant. The fitted **M**-dependent σ_{eq} model is shown as the red dashed curve; the model is also written in the top left of this figure. The long-dashed line indicates the imposed floor level of σ_{eq} .

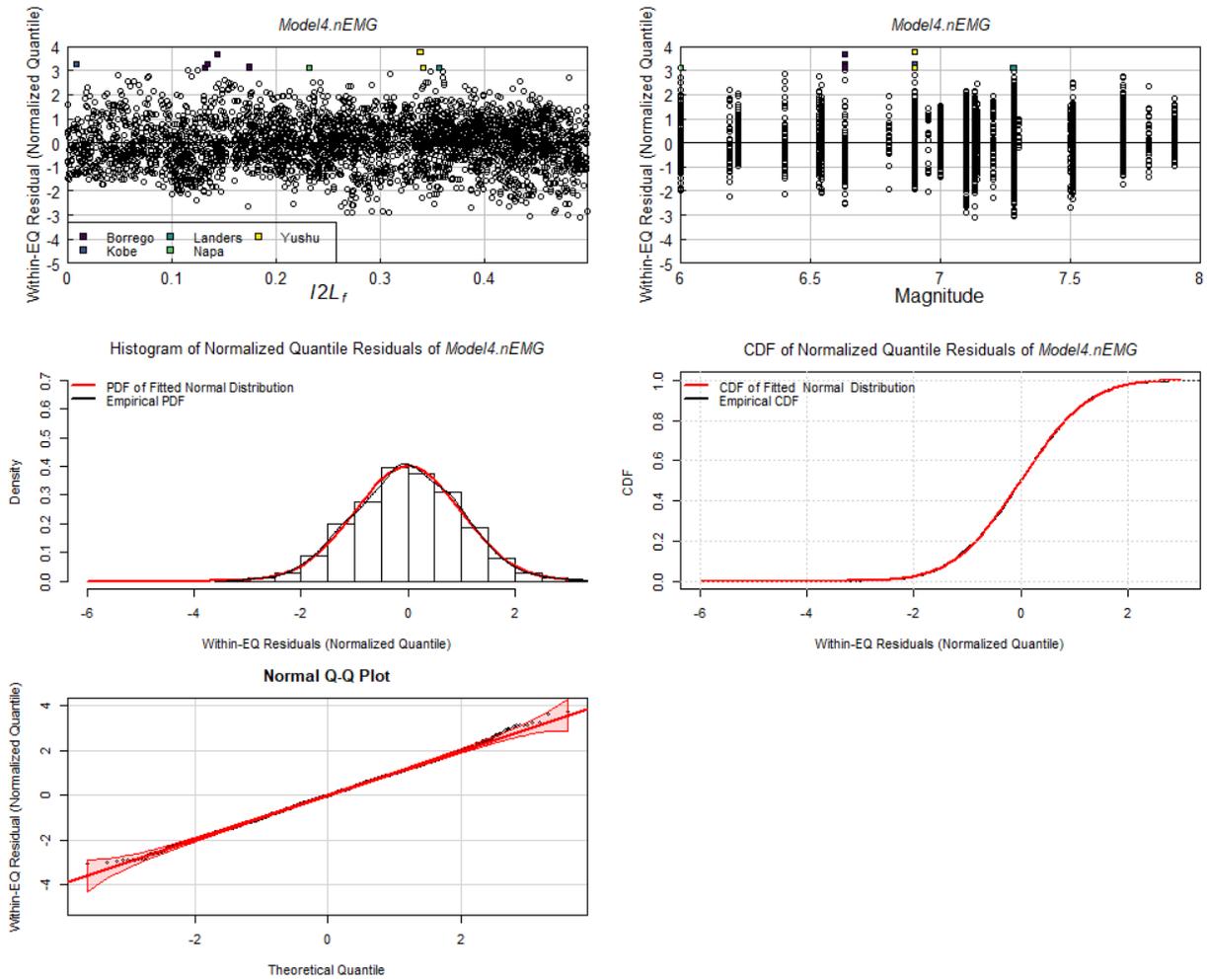


Figure 3.16. Residual diagnostic plots of *Model4.nEMG*. Within-earthquake (EQ) normalized quantile residuals are used on this figure. (Top left) residuals versus $l2L_f$. (Top right) residuals versus magnitude M . (Middle left) histogram of residuals; probability density function (PDF) of the normal distribution fitted to the residuals is plotted as the red curve. (Middle right) empirical cumulative distribution function (CDF) of residuals; CDF of the normal distribution fitted to the residuals is plotted as the red curve. (Bottom left) quantile-quantile (Q-Q) plot that compares the quantiles of normalized quantile residuals against the quantiles of standard normal distribution; the solid red line passes through the 1st and the 3rd quartiles, and the shaded area denotes roughly the 0.95 confidence level if residuals are taken from the standard normal distribution.

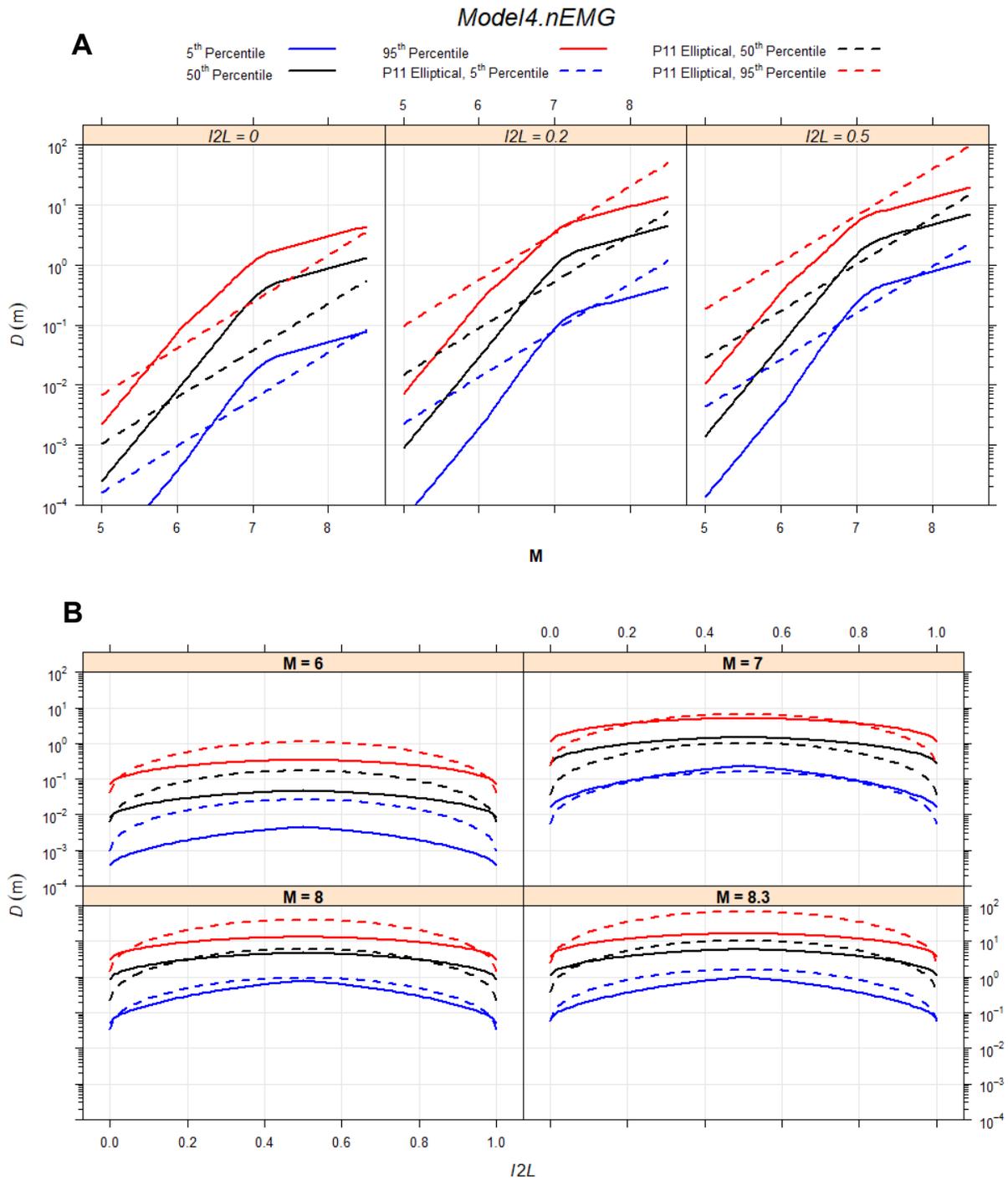


Figure 3.17. The 0.05, 0.50, and 0.95 quantiles (the 5th, 50th, and 95th percentiles in the figure title) of fault displacement (D) from the compound distribution of *Model4.nEMG*. (A) Quantiles versus magnitude M . (B) Quantiles versus normalized position along the main trace ($I2L$). For comparison, quantiles predicted by P11 (Petersen et al., 2011) are shown as the long-dashed curves.

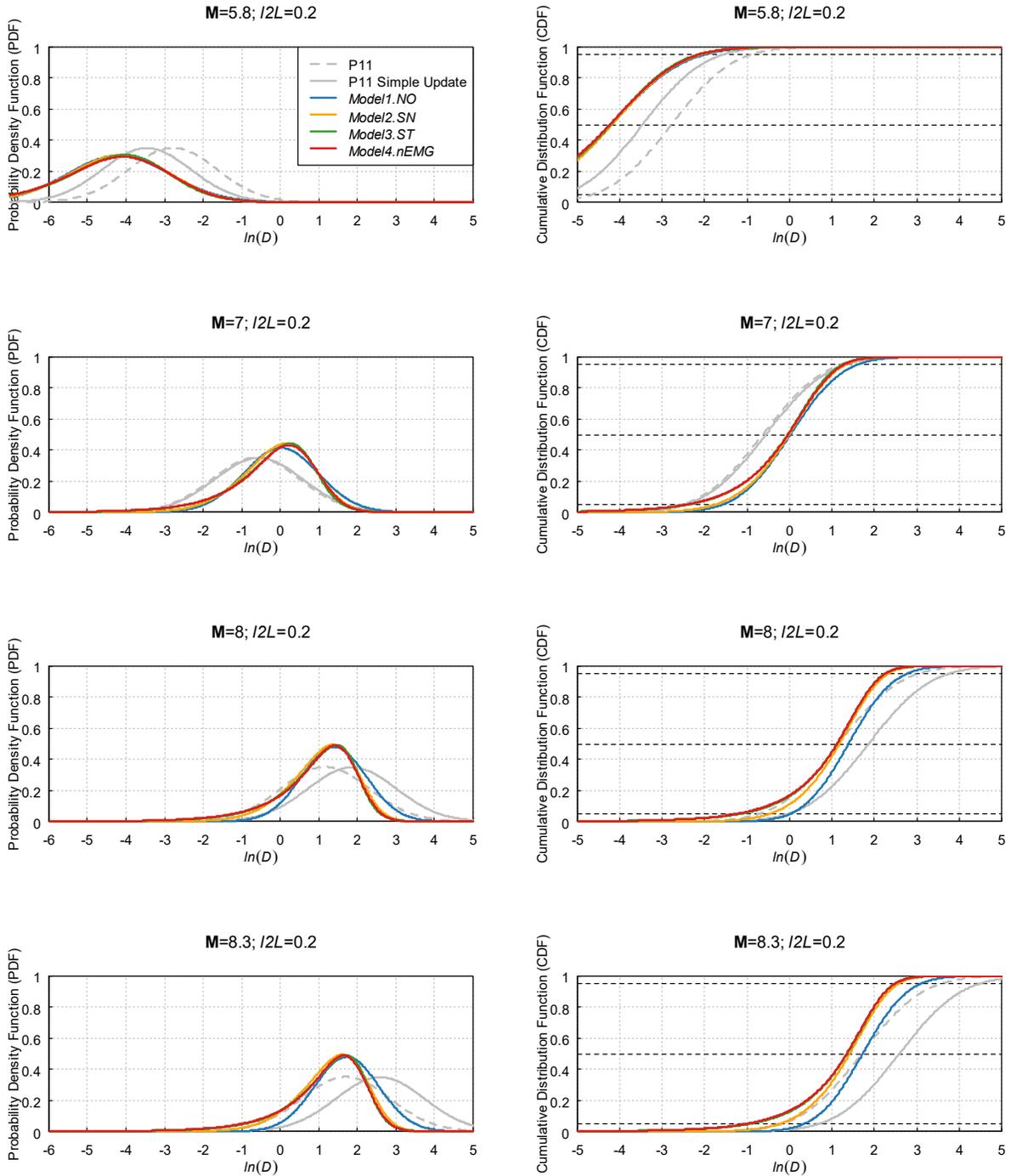


Figure 3.18. Comparison of compound distributions of fault displacement (D) from the models summarized in Table 3.1 and from Petersen et al. (2011) (P11). (Left) probability density functions (PDFs). (Right) cumulative distribution functions (CDFs) Calculations are carried out for ruptures of four different magnitudes (M), all at the same normalized position of $l2L = 0.2$.

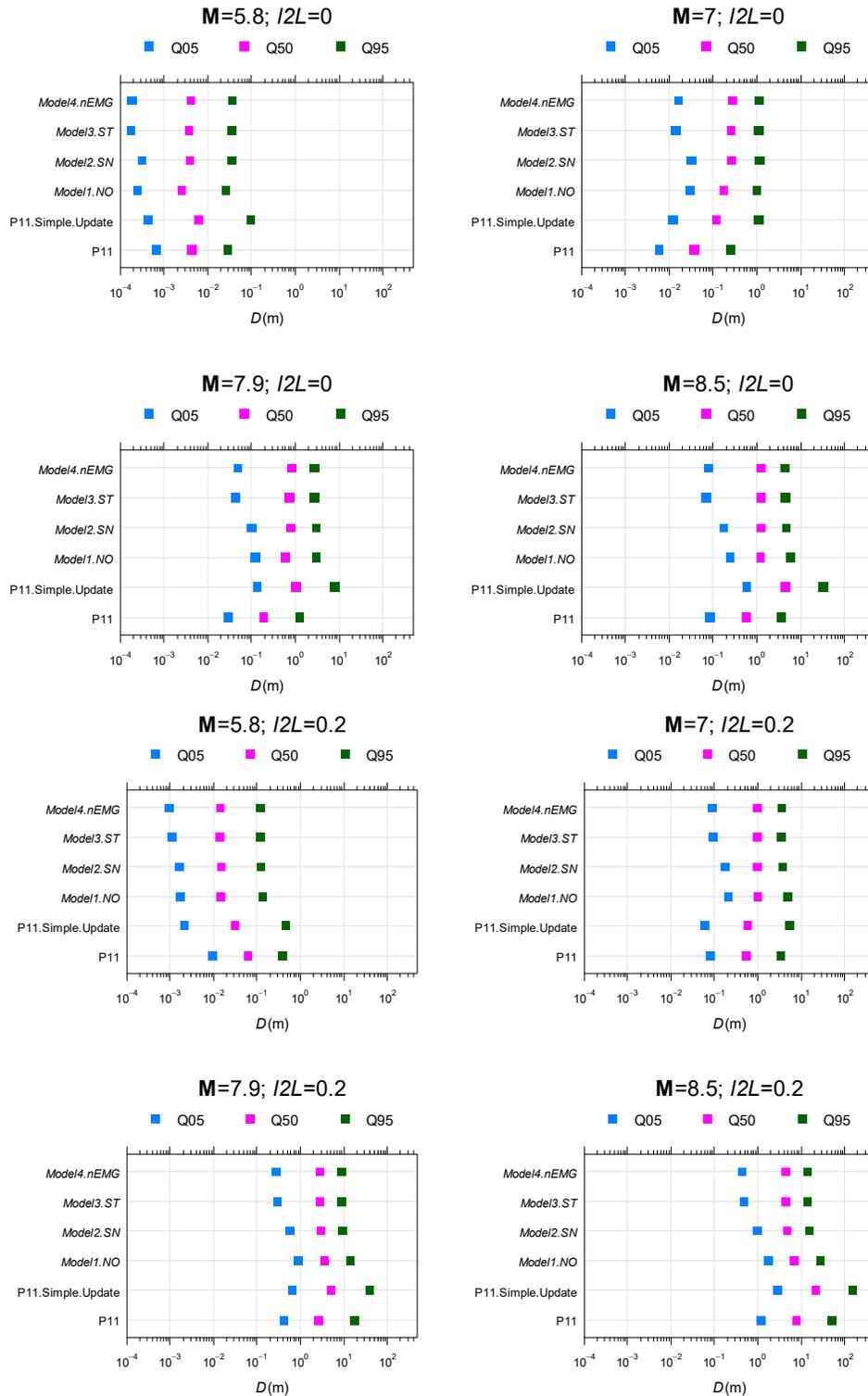


Figure 3.19. Comparison of the 0.05 (Q05), 0.50 (Q50), and 0.95 (Q95) quantiles of fault displacement (D) from the compound distributions of models summarized in Table 3.1 and from the Petersen et al. (2011) (P11). The rupture magnitude (M) and the site's normalized position ($l2L$) are indicated in the panel title.

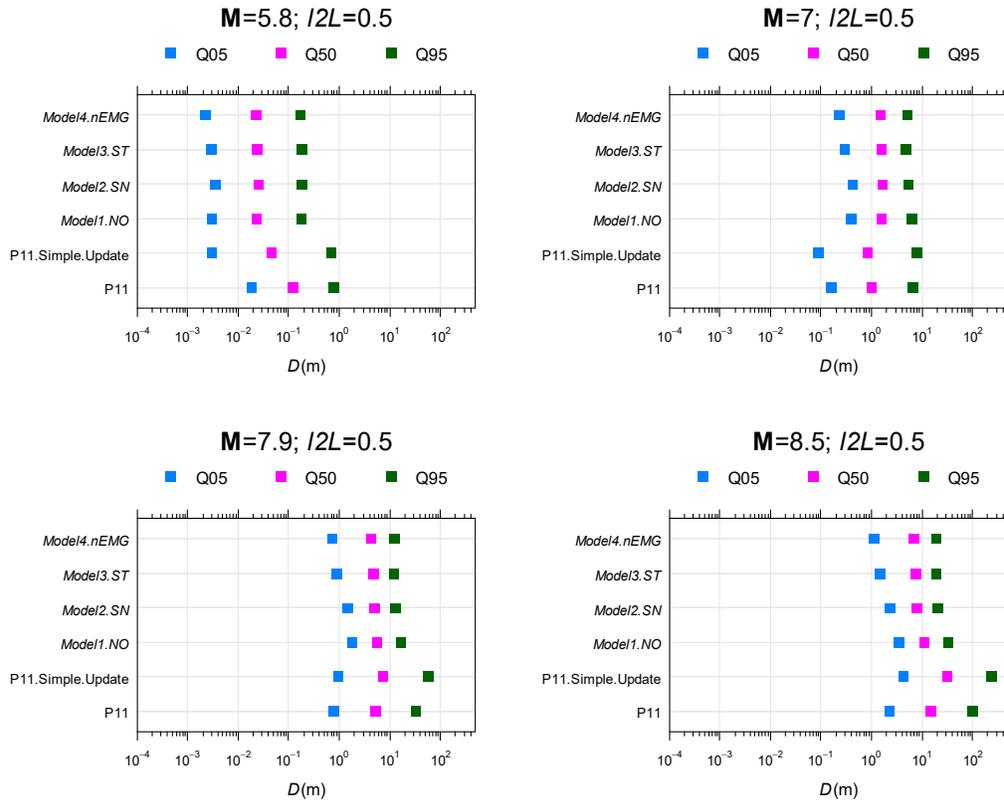


Figure 3.19. (Continued).

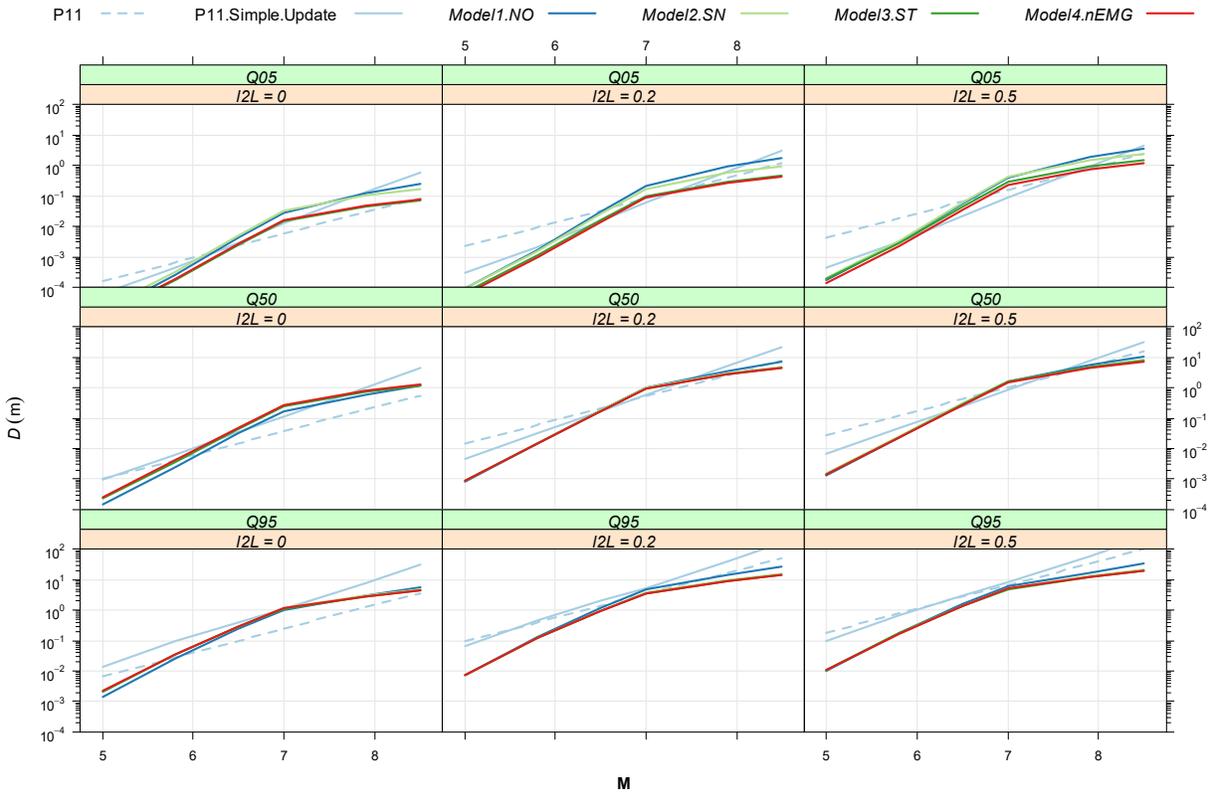


Figure 3.20. Quantiles of compound distribution of fault displacement (D) from the models summarized in Table 3.1 and from Petersen et al. (2011) (P11), plotted against magnitude (M). Each panel is conditional on the quantile ($Q05$, $Q50$, and $Q95$ for the 0.05-quantile, 0.5-quantile, and 0.95-quantile, respectively) and the normalized position along the main trace ($l2L = 0, 0.2$, and 0.5) as indicated in the strip of each panel.

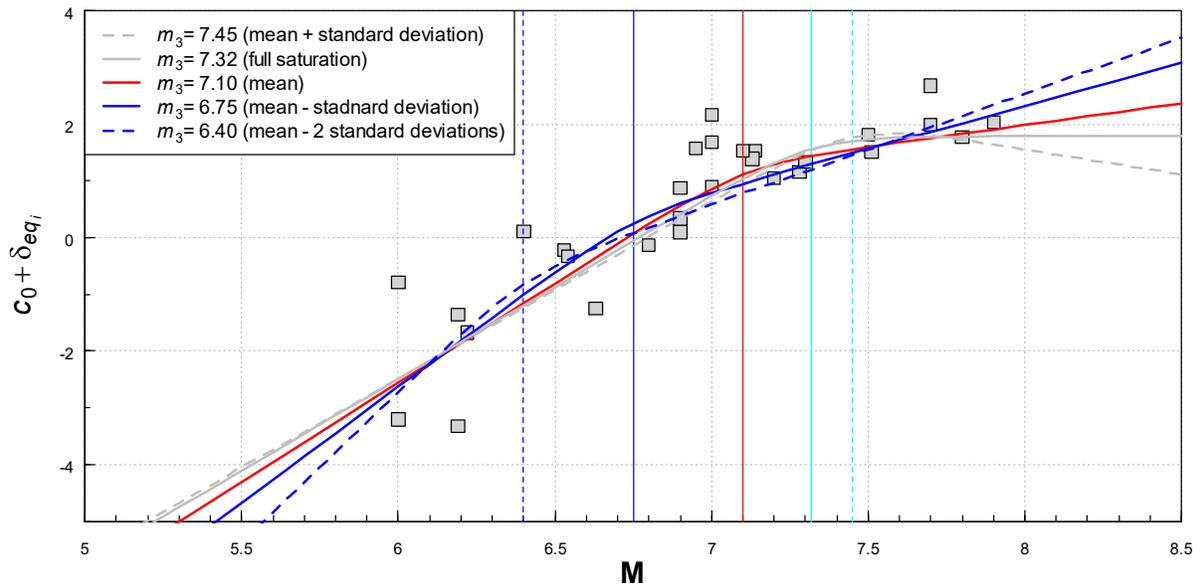


Figure 3.21. Magnitude (M)-scaling relations of fault displacement models (FDMs) regressed using an m_3 fixed to one of the five different values indicated in the plot legend. These m_3 values are marked in this figure by the color-coded vertical lines. For use as a reference for model comparison, the earthquake term ($c_0 + c_{eq,i}$) obtained from a regression analysis of an exploratory FDM without the magnitude scaling term f_M are plotted as the solid squares. Note that the five FDMs shown on this figure are based on the $\ln(D_{ij})$ data, not on the ($c_0 + c_{eq,i}$) values from the exploratory FDM.

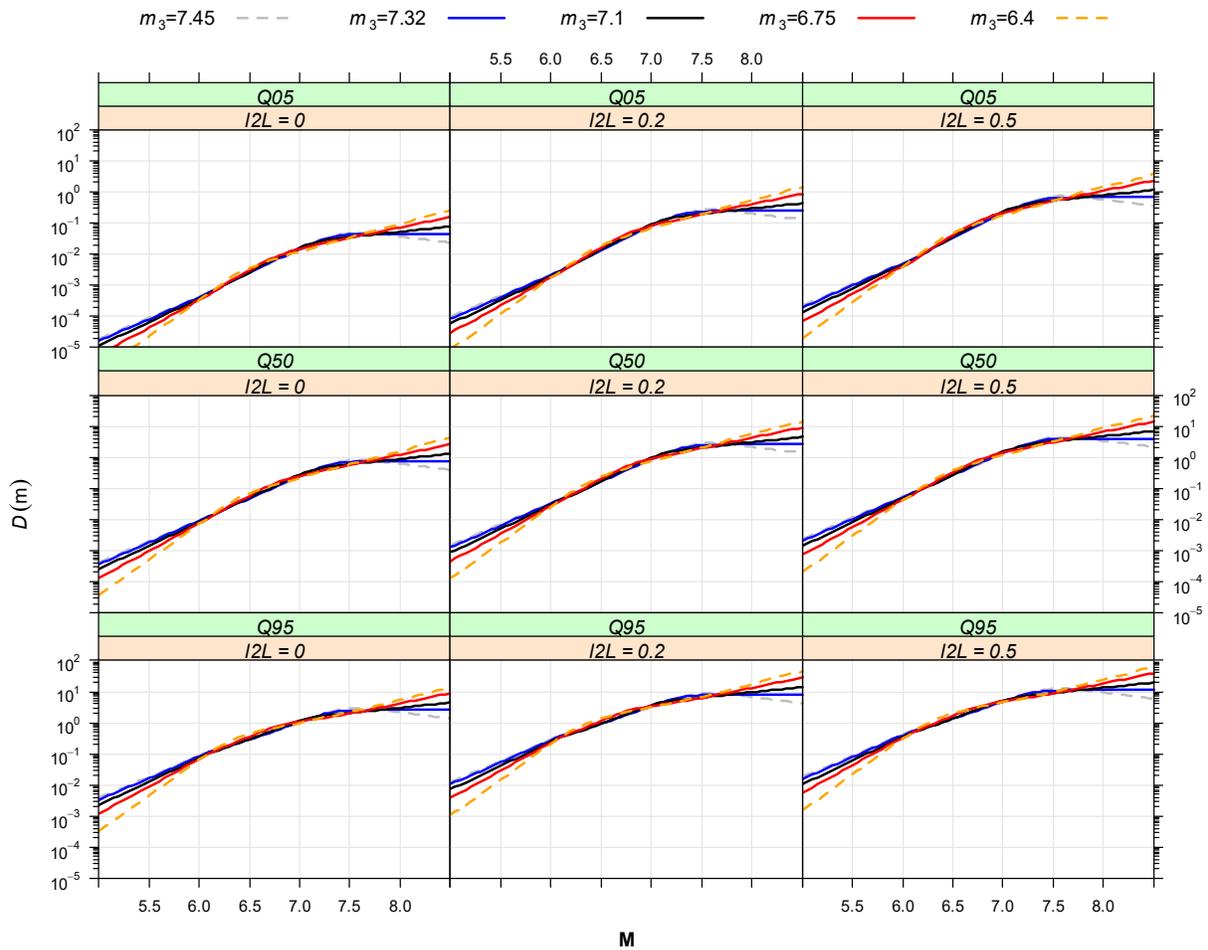


Figure 3.22. Quantiles of compound distributions of fault displacement (D) from the four models listed in Table 3.6, plotted versus magnitude (M). The fixed m_3 value used in each model is indicated in the figure legend. For reference, the quantiles from the model assuming $m_3 = 7.45$ are shown as the gray dashed curves. Each panel is conditional on the quantile (Q05, Q50, and Q95 for the 0.05-quantile, 0.5-quantile, and 0.95-quantile, respectively) and the normalized position along the main trace ($l2L = 0, 0.2,$ and 0.5) as indicated in the strip of each panel.

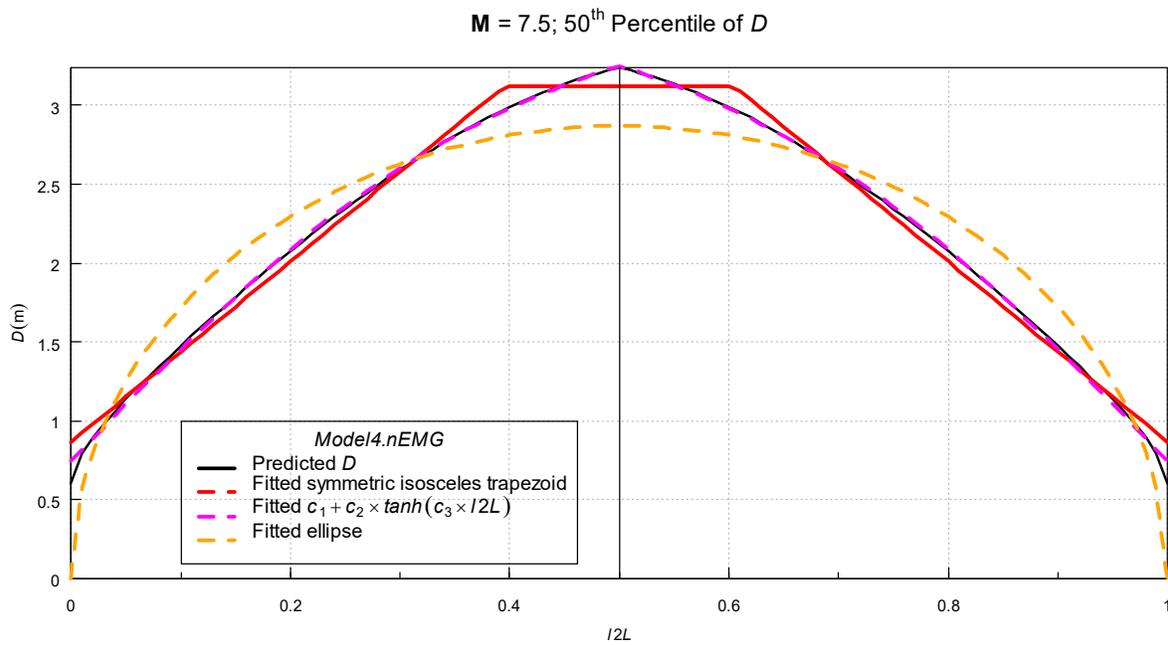


Figure 3.23. Profile of median slip predicted by *Model4.nEMG* for magnitude (**M**) 7.5 The predicted median is fitted by three functions of $l2L$ used in previous studies to characterize slip profile along the main trace. The fitted curves are shown as dashed curves in red, magenta, and orange color for the symmetric isosceles trapezoid, hyperbolic tangent (*tanh*) function, and ellipse equation, respectively.

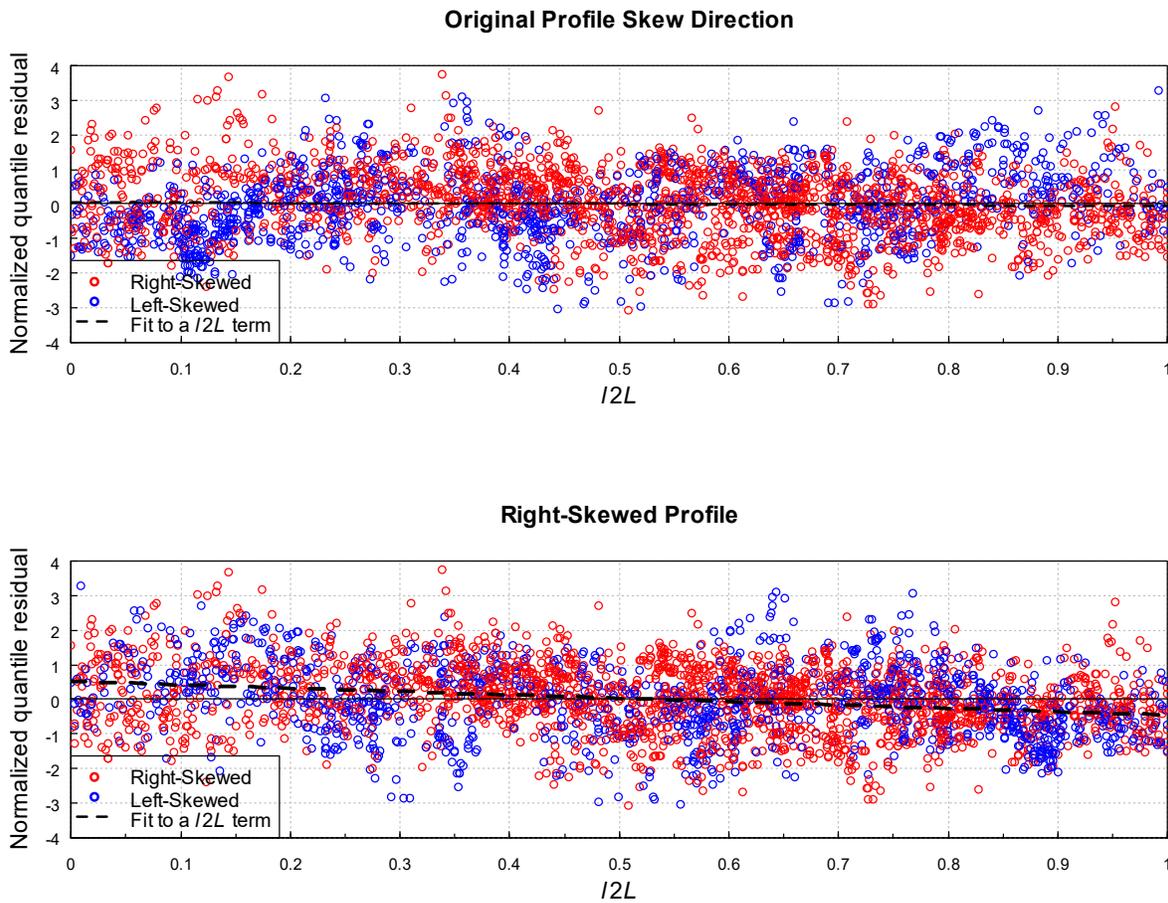


Figure 3.24. Plots of normalized quantile residuals from *Model4.nEMG*. (Top) The horizontal coordinate is the original normalized distance $l2L$ measured relative to the western end point of main trace. (Bottom) The horizontal coordinate is the normalized distance $l2L$ after left-skewed profiles are reflected. A linear function of $l2L$ is fitted to the residuals in each panel and the fitted curves are shown as the dashed lines.

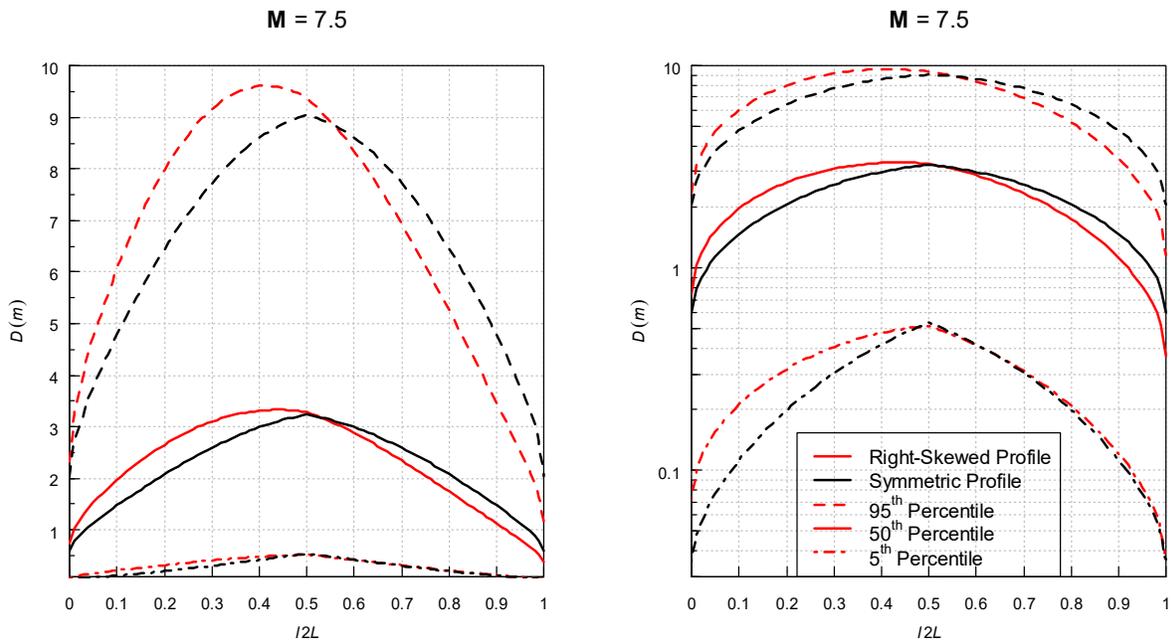


Figure 3.25. Predicted quantiles of displacement (D) versus normalized position along the main trace ($l2L$), for a magnitude (M) 7.5 rupture. Black curves are computed from *Mode4.nEMG*, a spatially symmetric model. Red curves are computed from the spatially asymmetric *Mode6.nEMG*. Note that the peak of predicted quantile from *Mode6.nEMG* occurs near $l2L = 0.4$.

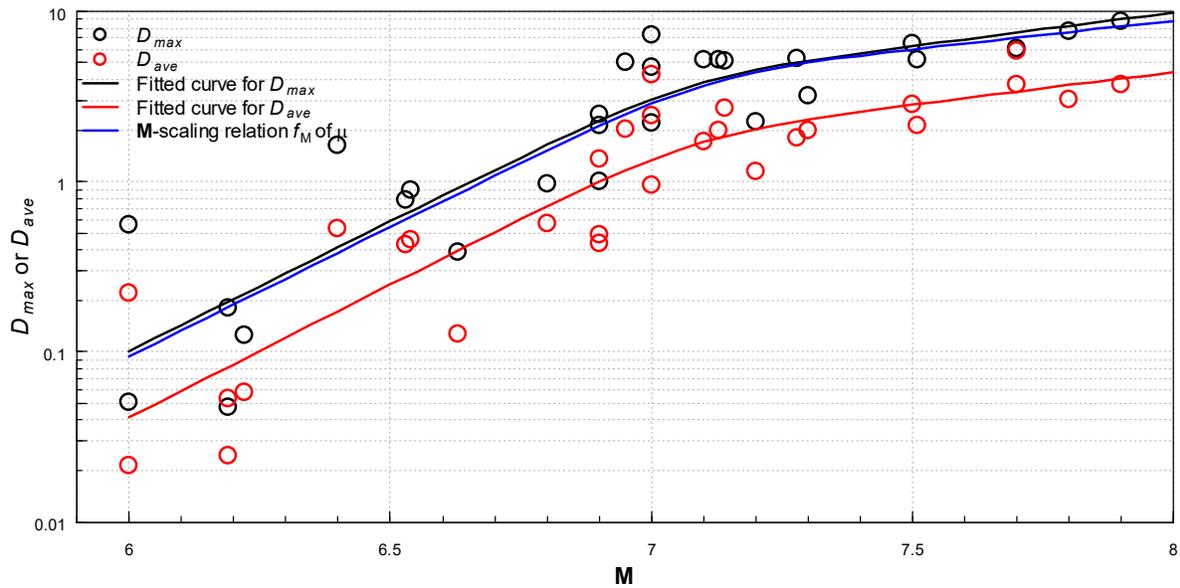


Figure 3.26. The average (D_{ave}) and the maximum (D_{max}) of observed displacements for each strike-slip earthquake selected from the Fault Displacement Hazard Initiative (FDHI) database, plotted against magnitude (M). The fits by the bilinear function of M (Equations (3.23) and (3.24)) are shown as the red curve and the black curve for D_{ave} and D_{max} , respectively. For reference, the M -scaling relation $c_0 + f_M$ of *Model4.nEMG* is shown as the blue curve.

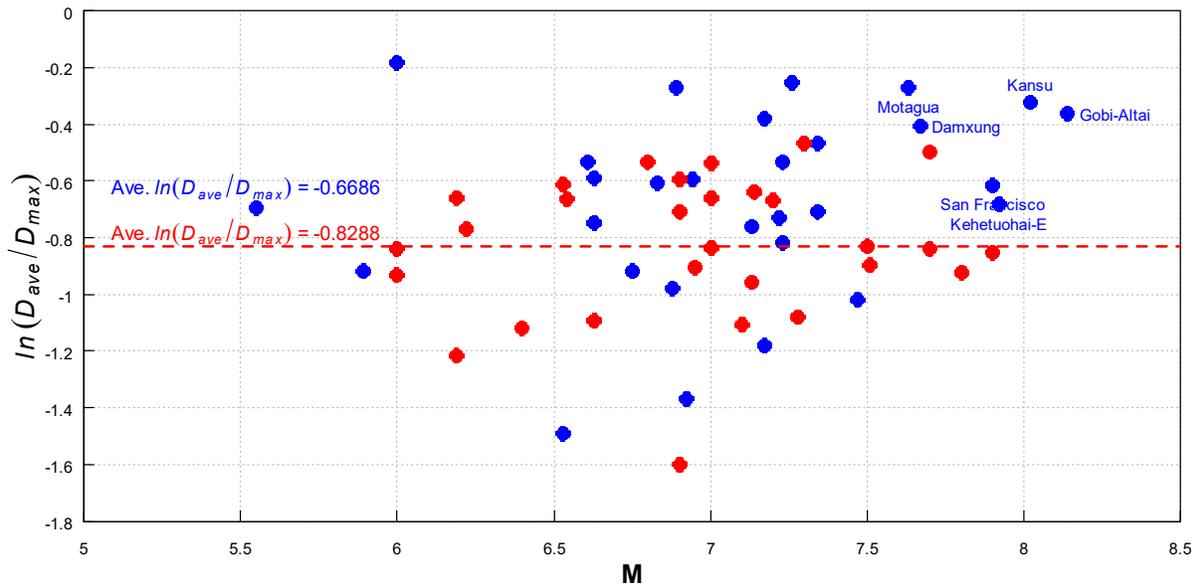


Figure 3.27. Logarithm of the ratio D_{ave}/D_{max} , plotted against magnitude (M). The average (D_{ave}) and the maximum (D_{max}) of observed displacements in an individual earthquake are obtained from two data sources. The blue circles are from the Fault Displacement Hazard Initiative (FDHI) database (Sarmiento et al., 2021), and the red circles are from Wells and Coppersmith (1994). The average value of $\ln(D_{ave}/D_{max})$ from each dataset is marked by the horizontal dashed line.

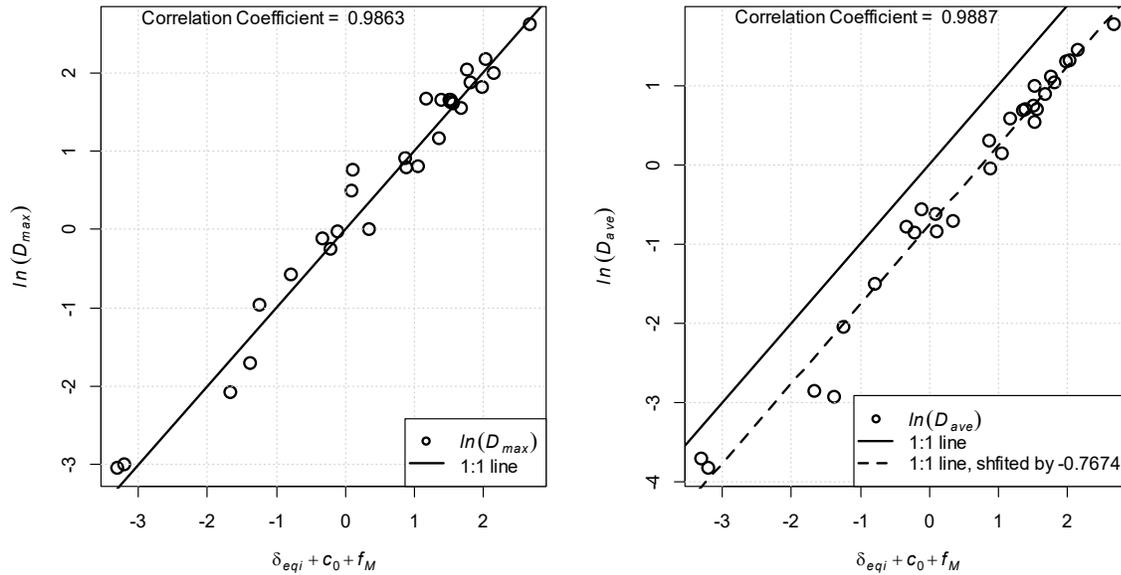


Figure 3.28. (left) Logarithm of observed maximum displacement ($\ln(D_{max})$) versus the magnitude-scaling term ($\delta_{eq,i} + c_0 + f_M$) of *Model4.nEMG*. (right) Logarithm of observed average displacement ($\ln(D_{ave})$) versus ($\delta_{eq,i} + c_0 + f_M$) of *Model4.nEMG*. Both $\ln(D_{max})$ and $\ln(D_{ave})$ have a 1:1 relation with the ($\delta_{eq,i} + c_0 + f_M$) term. The standard deviation of the scatter around the 1:1 line is 0.2548 and 0.2353 for $\ln(D_{max})$ and $\ln(D_{ave})$, respectively.

M = 7.2

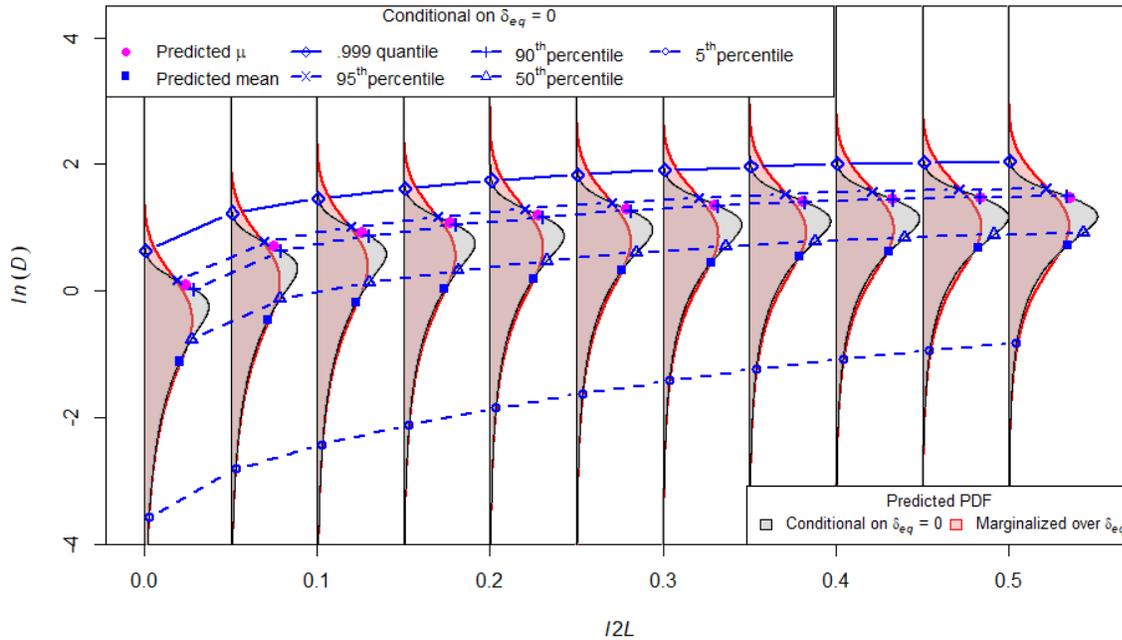


Figure 3.29. Predicted probability density function (PDF) of fault displacement (D) versus the normalized position along the main trace ($l2L$) of an magnitude (M) 7.2 rupture. The predicted PDF conditional on zero random intercept ($\delta_{eq,i} = 0$) is shown by the gray curve. The predicted mean of the Gaussian component (μ) (magenta square) falls between the predicted 90th and 95th percentiles of displacements. The predicted mean displacement ($\mu - v$, where v is the mean of the Exponential component) (blue square) falls below the 50th percentile, as expected for a left-skewed distribution. For reference, the PDF of the compound distribution resulting from the marginalization over δ_{eq} is shown as the red curve.

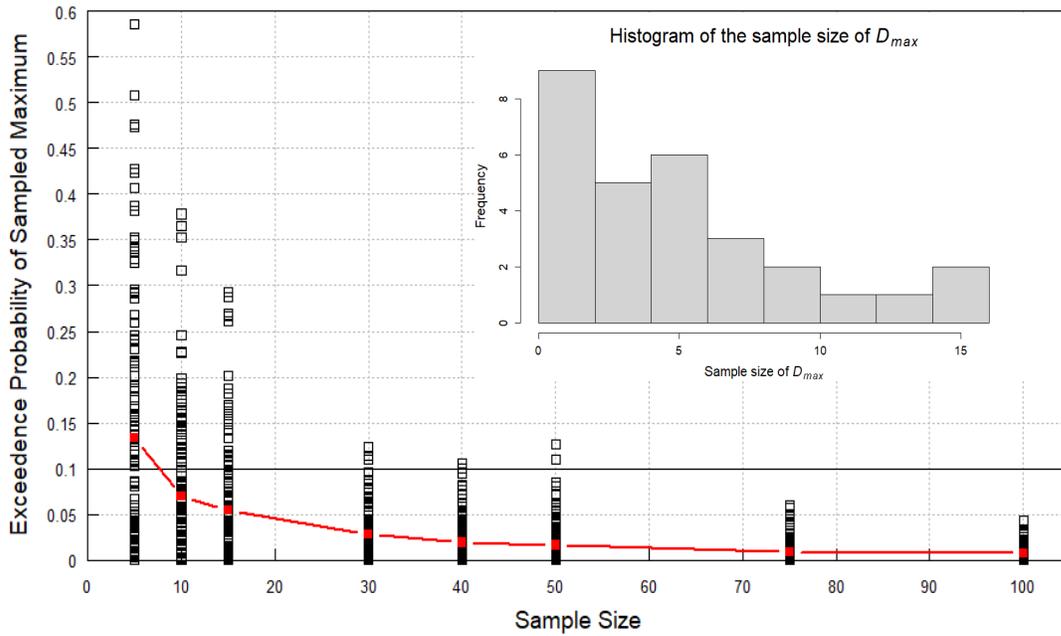


Figure 3.30. Simulated exceedence probability of sample maximum, plotted against sample size. Samples are taken from the predicted nEMG distribution for magnitude (M) 7.3 at the normalized along-trace location of $l2L = 0.5$, for eight different sample sizes. To account for sampling variability, sampling is repeated 100 times for each sample size. The red solid square marks the mean of exceedence probability over these 100 trials. (Inset) Histogram of the sample size of displacements, per earthquake, in Fault Displacement Hazard Initiative (FDHI) database that fall within a 1-km-wide spatial window centered at the location of observed maximum displacement (D_{max}).

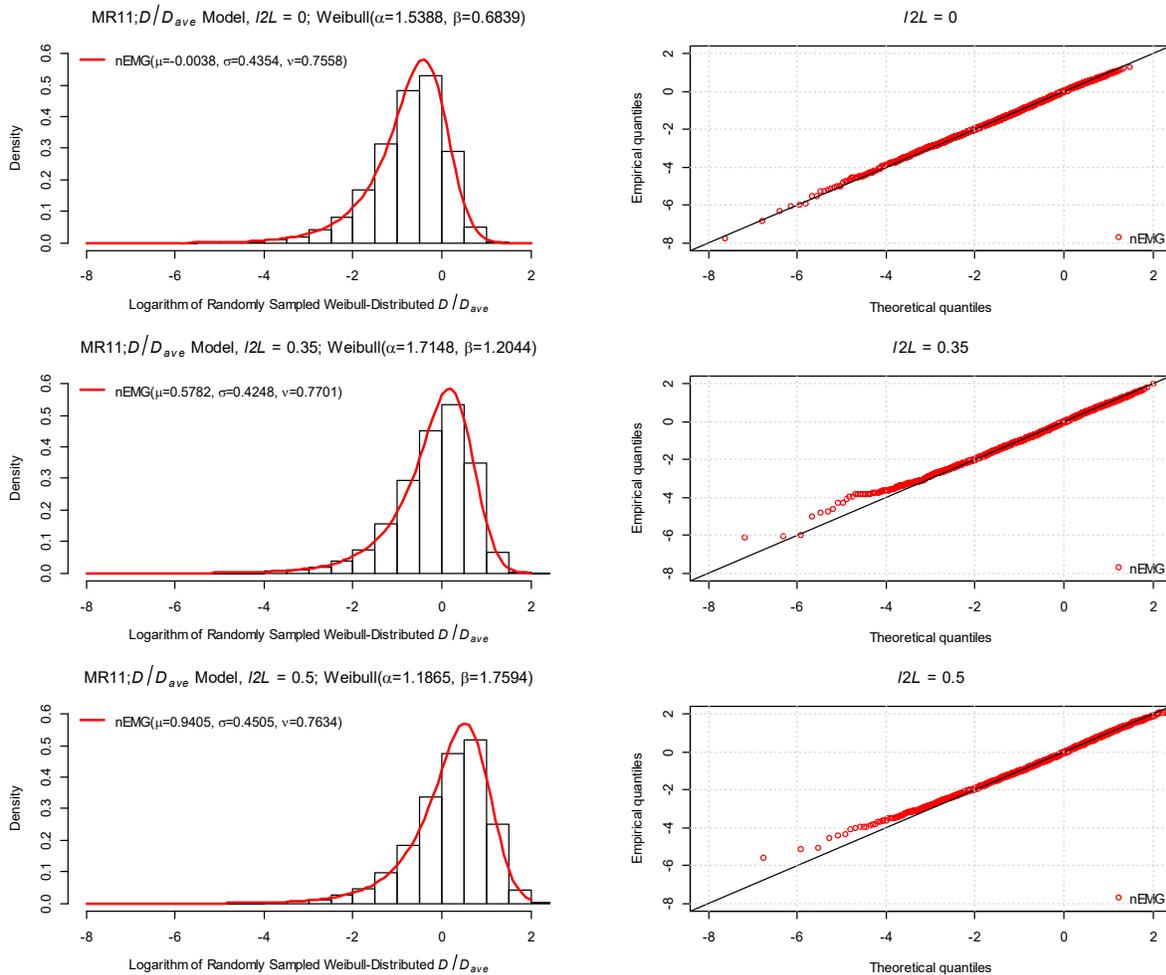


Figure 3.31. Matching the logarithm of Weibull-distributed ratio of displacement (D) to average displacement (D_{ave}), D/D_{ave} , by the negative exponentially modified Gaussian (nEMG) distribution. (Left) Histogram of the logarithm of D/D_{ave} sampled from a Weibull distribution whose parameters, as shown in the panel title, are predicted by the D/D_{ave} model of Moss and Ross (2011; MR11). The density function of the matched nEMG distribution is shown as the red curve. Parameters of the matched nEMG distribution are given in the plot legend. (Right) Quantile-quantile plot that compares quantiles of the logarithm of sampled D/D_{ave} versus the theoretical quantile of the matched nEMG distribution.

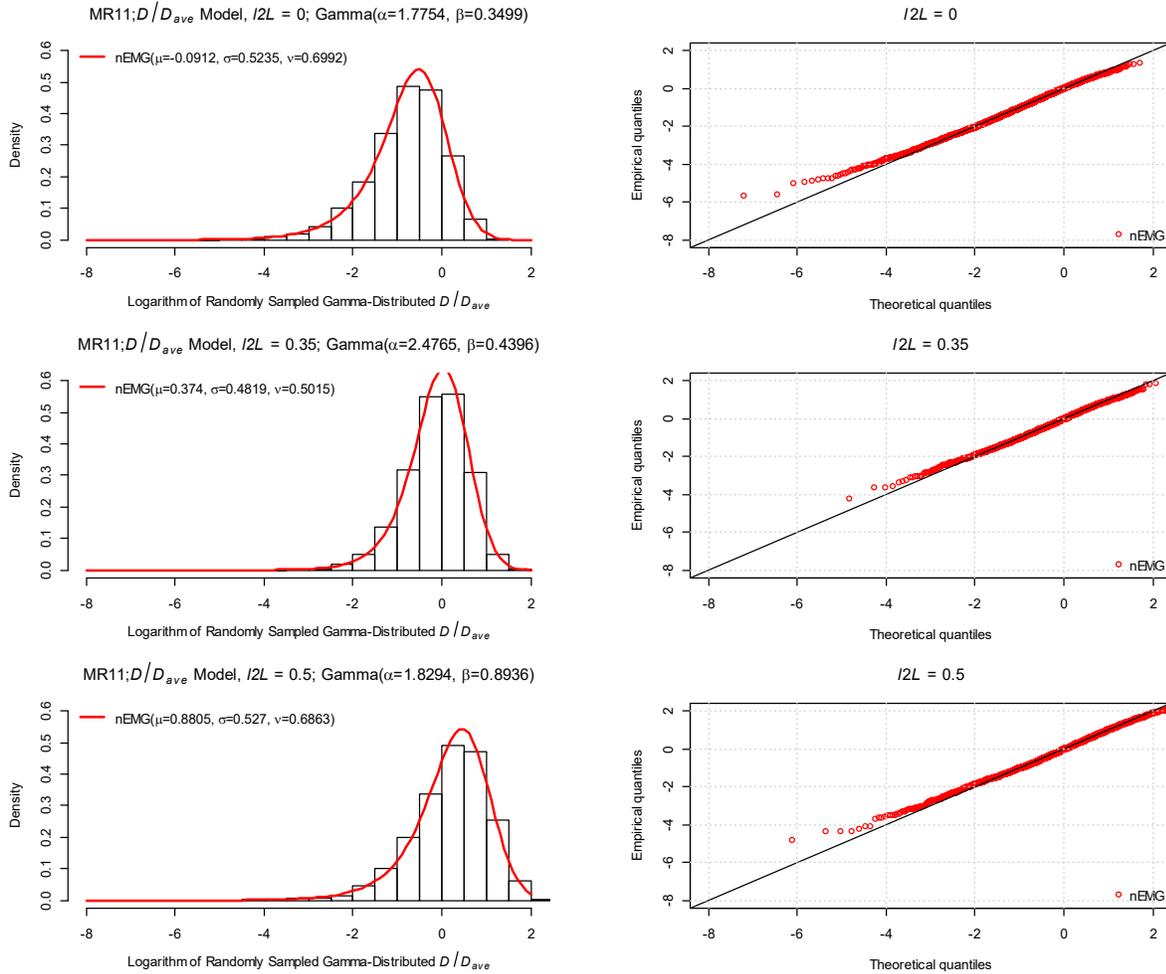


Figure 3.32. Matching the logarithm of gamma-distributed ratio of displacement (D) to average displacement (D_{ave}), D/D_{ave} , by the negative exponentially modified Gaussian (nEMG) distribution. (Left) Histogram of the logarithm of D/D_{ave} sampled from a gamma distribution whose parameters, as shown in the panel title, are predicted by the D/D_{ave} model of Moss and Ross (2011; MR11). The density function of the matched nEMG distribution is shown as the red curve. Parameters of the matched nEMG distribution are given in the plot legend. (Right) Quantile-quantile plot that compares quantiles of the logarithm of sampled D/D_{ave} versus the theoretical quantile of the matched nEMG distribution.

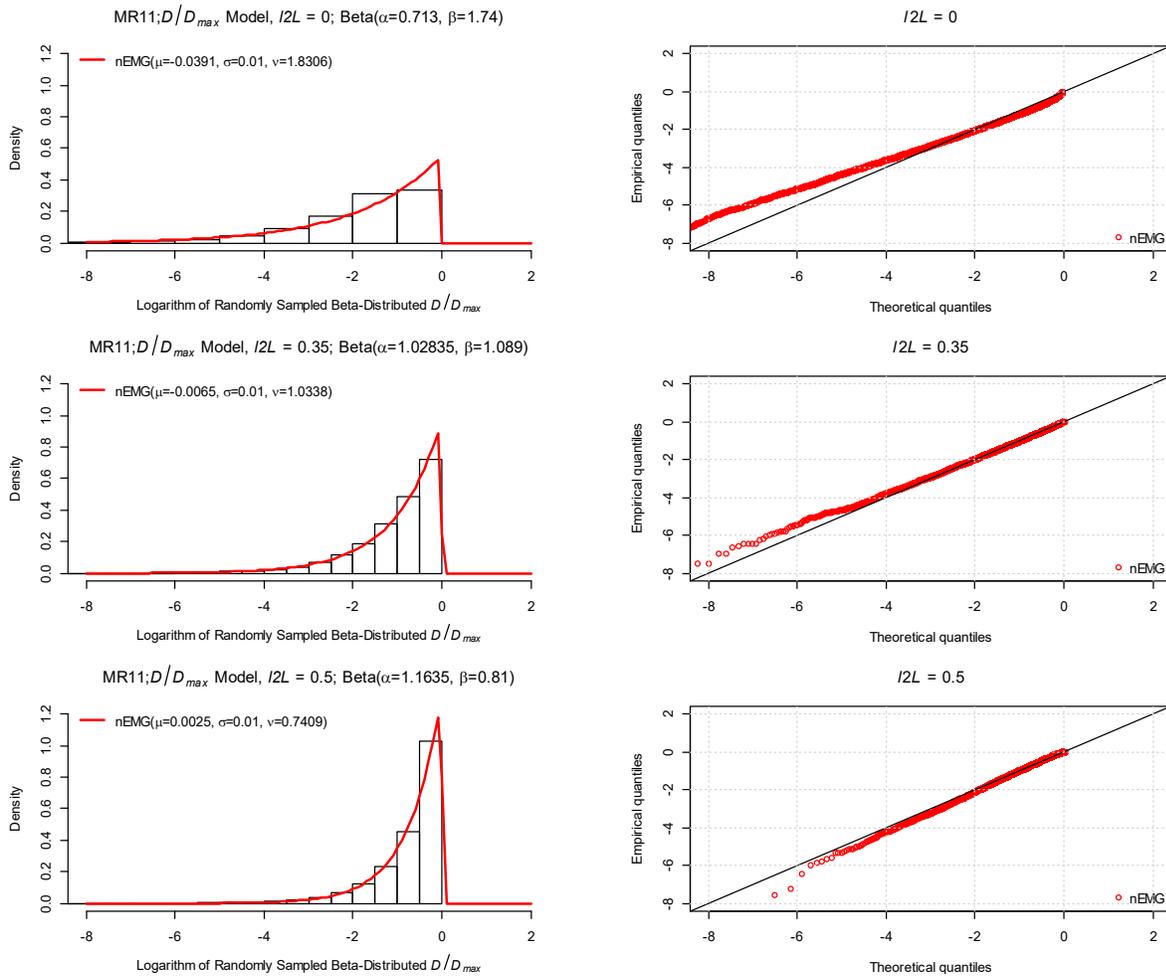


Figure 3.33. Matching the logarithm of beta-distributed ratio of displacement (D) to average displacement (D_{max}), D/D_{max} , by the negative exponentially modified Gaussian (nEMG) distribution. (Left) Histogram of the logarithm of D/D_{max} sampled from a beta distribution whose parameters, as shown in the panel title, are predicted by the D/D_{max} model of Moss and Ross (2011; MR11). The density function of the matched nEMG distribution is shown as the red curve. Parameters of the matched nEMG distribution are given in the plot legend. (Right) Quantile-quantile plot that compares quantiles of the logarithm of sampled D/D_{max} versus the theoretical quantile of the matched nEMG distribution.

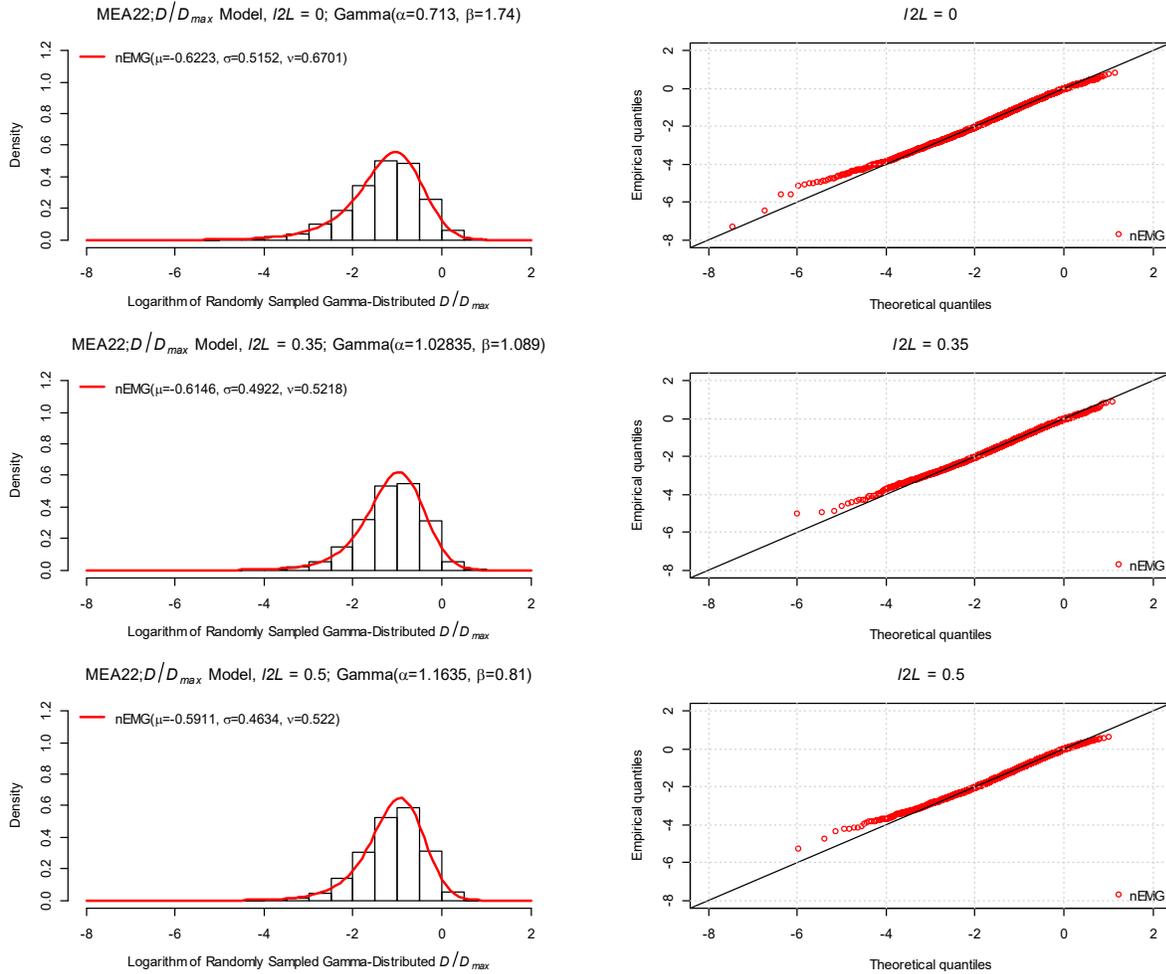


Figure 3.34. Matching the logarithm of gamma-distributed ratio of displacement (D) to maximum displacement (D_{max}), D/D_{max} by the negative exponentially modified Gaussian (nEMG) distribution. (Left) Histogram of the logarithm of D/D_{max} sampled from a gamma distribution whose parameters, as shown in the panel title, are predicted by the D/D_{max} model of Moss et al. (2022; MEA22). The density function of the matched nEMG distribution is shown as the red curve. Parameters of the matched nEMG distribution are given in the plot legend. (Right) Quantile-quantile plot that compares quantiles of the logarithm of sampled D/D_{max} versus the theoretical quantile of the matched nEMG distribution.

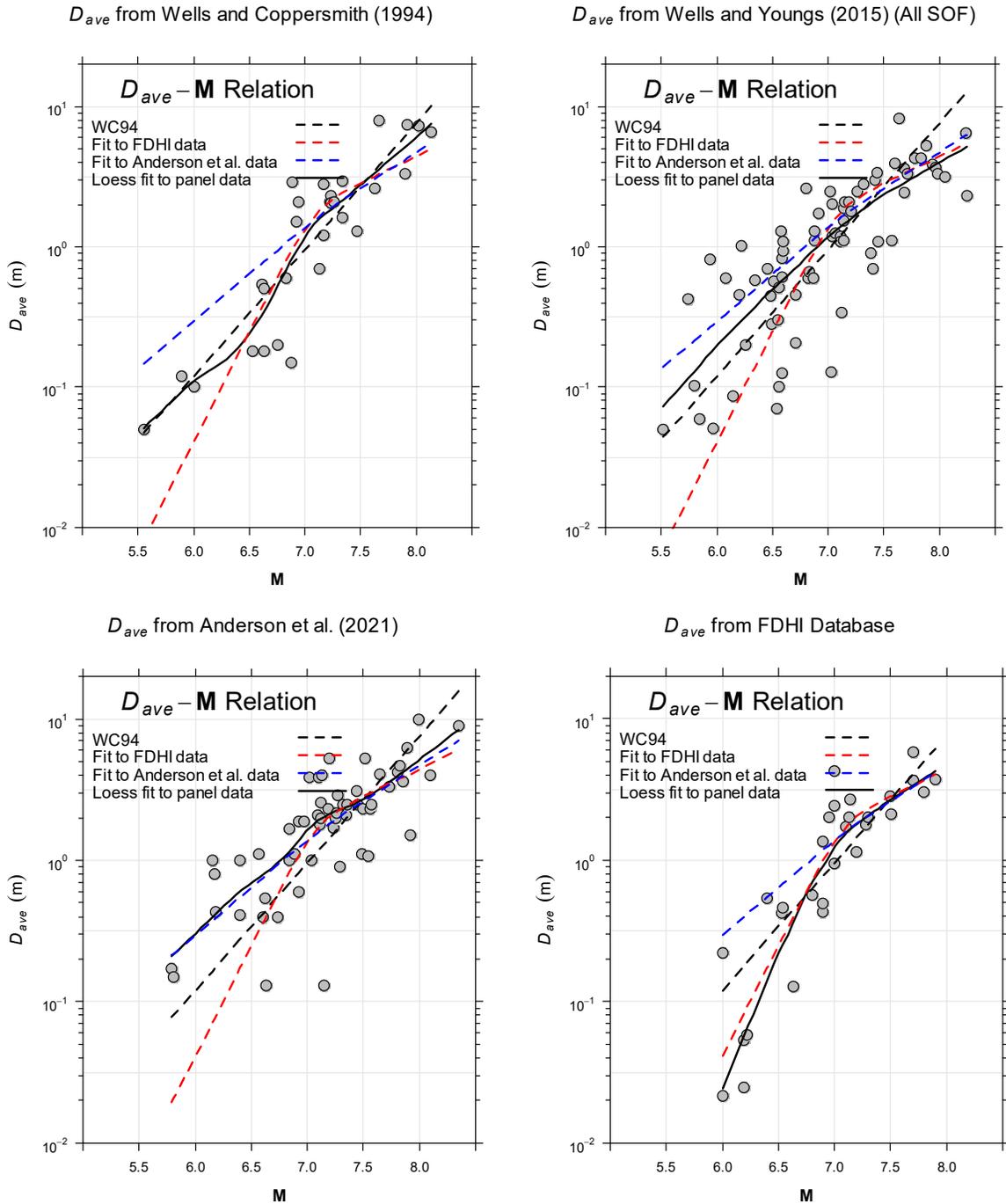
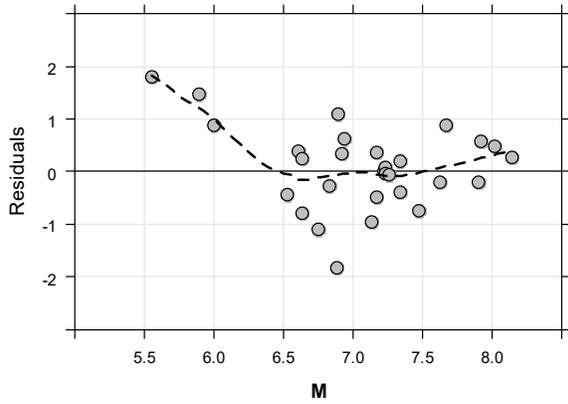
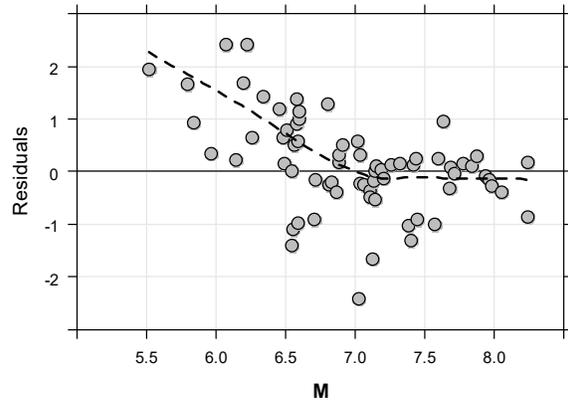


Figure 3.35. Scatter plots of observed average displacement (D_{ave}) versus magnitude (M). (Top left) data from Wells and Coppersmith (1994; WC94); (Top right) data from Wells and Youngs (2015; WY15); all styles of faulting (SOF); (Bottom left) data from Anderson et al. (2021; AEL21); (Bottom right) Fault Displacement Hazard Initiative (FDHI) database (Sarmiento et al., 2021) used in this study. A *loess* smooth is calculated and plotted as solid black curve in each panel. For reference, three parametric models fitted to the datasets of WC94, FDHI, and AEL21 are shown as dashed curves in black, red, and blue color, respectively.

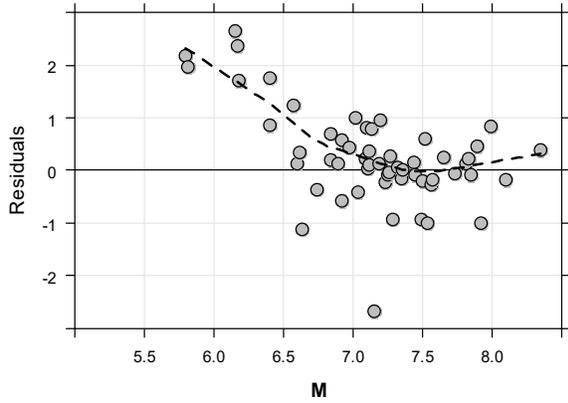
Residuals of D_{ave} data from Wells and Coppersmith (1994)



Residuals of D_{ave} data from Wells and Youngs (2015)



Residuals of D_{ave} data from Anderson et al. (2021)



Residuals of D_{ave} data from FDHI database

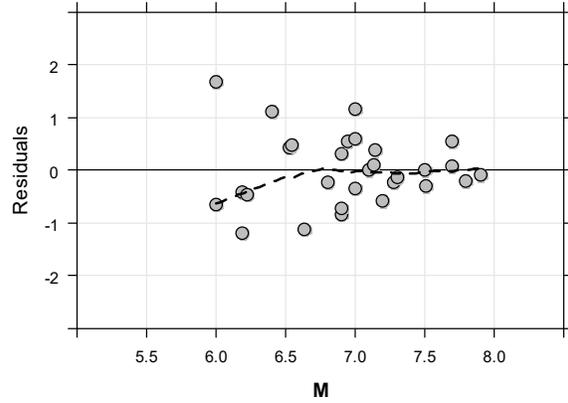


Figure 3.36. Data residuals with respect to the D_{ave} - magnitude (M) relation of Equation (3.23) developed using the average displacement (D_{ave}) data from the Fault Displacement Hazard Initiative (FDHI) database (Sarmiento et al., 2021). To help visualize the residual trend, the smooth curve fitted by the nonparametric, locally weighted (*loess*) regression on the data residuals is shown as the dashed curve.

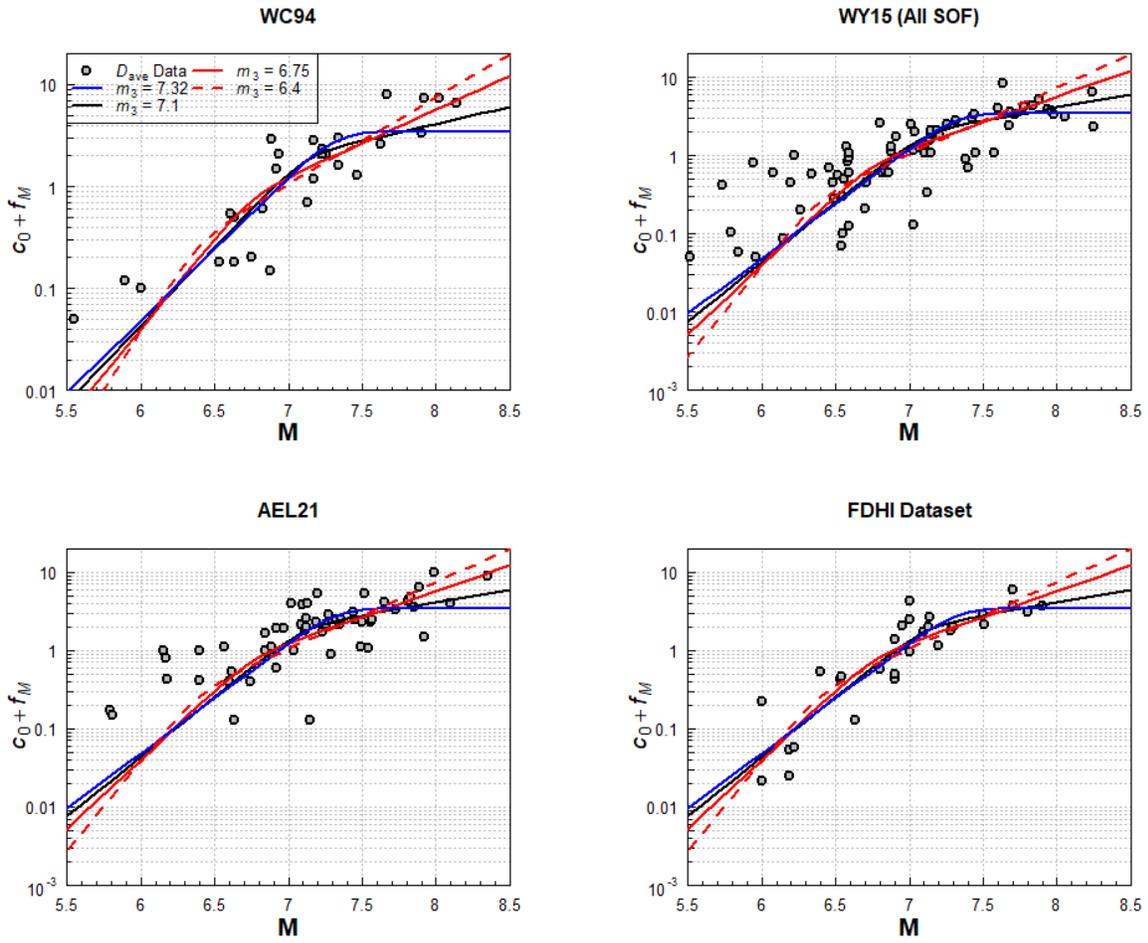


Figure 3.37. The magnitude-scaling terms ($c_0 + f_M$) from *Model4.nEMG* and its three epistemic variants, all vertically shifted by -0.7642 log units, are plotted against magnitude (M). For comparison, average displacement (D_{ave}) data from four different data sources (Wells and Coppersmith (1994; WC94); Wells and Youngs (2015; WY15); Anderson et al. (2021; AEL21) and Fault Displacement Hazard Initiative (FDHI) database (Sarmiento et al., 2021)) are plotted in four separate panels as the solid gray symbols. WY15 data include strike-slip as well as other styles of faulting (SOF).

4 Example Hazard Applications

We adopt the PFDHA framework established in P11. Hazard equations and component models are summarized in Appendix A. This chapter presents numerical implementations of hazard integrations and discusses how the new FDMs fit in the overall framework. We demonstrate the use of the new FDMs in hazard applications, including the sensitivity of hazard to epistemic uncertainties. Hazard curves from our preferred model, *Model4.nEMG*, are compared with those using the P11 model and the interim models described in Section 3.2.

4.1 NUMERICAL IMPLEMENTATION FOR SCENARIO EVENTS

In both PSHA and PFDHA, hazard integrations often are carried out numerically. For the example applications in this Chapter, we assume a given fault ruptures repetitively with similar magnitudes (i.e., the characteristic earthquake recurrence model), or the occurrence of earthquakes can be characterized by a catalog of scenario earthquakes, each of which has a specified magnitude, rupture location on the fault, and occurrence rate. Hazard equation for principal displacement, Equation (A.1) in Appendix A, can then be written in the summation form as:

$$\begin{aligned} & \lambda(D \geq D_0) \\ &= \sum_{i=1}^{N_m} \alpha(m_i) P[sr \neq 0 | m_i] P_2[D \neq 0 | \Delta_{site,i}, z, sr \neq 0] P_1[D \neq 0 | z, sr \neq 0] P[D \geq D_0 | (l/L)_i, m_i, D \neq 0] \end{aligned} \quad (4.1)$$

where N_m is the total number of earthquakes in the catalog, m_i is magnitude of the i^{th} earthquake, and $\alpha(m_i)$ is the annual rate of the i^{th} earthquake. See Figure 2.1 and Appendix A for other parameters.

The exceedance probability, $P[D \geq D_0 | (l/L)_i, m_i, D \neq 0]$, is complementary of the CDF of the assumed probability distribution model for the principal displacement. Using the FDM of P11 as an example, the natural log of fault displacement, $\ln(D)$, is assumed to be normally distributed. The normal probability density function is

$$N(y; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} \quad (4.2)$$

where $y = \ln(D)$, μ is the mean, and σ is the standard deviation of the normal distribution.

Exceedance probability is

$$P(Y > y_0) = 1 - \int_{-\infty}^{y_0} N(y; \mu, \sigma) dy = 1 - \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{y_0} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} dy \quad (4.3)$$

In our PFDHA FORTRAN code, this integration is achieved using FORTRAN's built-in error function, which also works for the interim model, *Modell.NO*.

PDFs and CDFs for the skew-normal, skew- t , and nEMG distributions are presented, and their implementations in R and in FORTRAN are briefly explained in Appendix C. The skew-normal and the nEMG distributions have been implemented in our PFDHA code and used in hazard calculations for the example applications described in Section 4.2. FORTRAN implementation for the skew- t distribution is yet to be incorporated into our PFDHA code. The FORTRAN source codes for PDF and CDF are not listed in Appendix C, but they can be downloaded from <https://www.conservation.ca.gov/cgs/pfdha>. FORTRAN code for PFDHA will be provided upon request.

4.2 EXAMPLE APPLICATIONS

Three simple scenario example applications are illustrated: the **M** 7.0 scenario used in P11, an **M** 5.8 scenario used in the IAEA benchmarking study (Valentini et al., 2021), and a hypothetical **M** 8.3 scenario. We also present hazard curves for a case with a seismic source logic tree developed by the IAEA for its benchmarking study (Valentini et al., 2021).

For illustrating the application of the new FDMs presented in Chapter 3, we limit the example sites to be on the mapped fault ($\Delta_{site,i} = 0$), and location uncertainty is not considered ($P_2[D \neq 0 | \Delta_{site,i}, z, sr \neq 0] = 1$). Following the P11 implementation, the principal rupture is assumed to occur everywhere along the main trace of earthquake i (i.e., the term $P_1[D \neq 0 | z, sr \neq 0]$ is 1). For all examples, only principal displacement is analyzed, and the Wells and Coppersmith (1993) probability of surface rupture, Equation (A.2) in Appendix A, is applied.

In the following, hazard curves for an on-fault site using our preferred FDM, *Model4.nEMG*, are presented and compared with those using the P11 model and two other interim models, *Modell.NO* and *Model2.SN*, presented in Chapter 3.

4.2.1 M 7.0 scenario example in Petersen et al. (2011)

The scenario example in P11 was developed based on Hayward fault geometry and paleoseismic information. It was assumed that the fault ruptures repetitively with an **M** 7.0 earthquake occurring once every 140 years (annual rate of 7.14×10^{-3}) and the site was located on the fault at $l/L = 0.4055$.

In Figure 4.1, hazard curves calculated using the preferred model, *Model4.nEMG*, the P11 bilinear (with respect to l/L) model, and two interim models (*Model1.NO* and *Model2.SN*) are compared. The new models predict larger displacement compared to the P11 bilinear model except at exceedance rate smaller than about 10^{-4} , largely because the new bilinear **M**-scaling yields higher predicted displacement values for magnitudes around 7 as shown in Figure 3.7. The preferred model and the skew normal model (*Model2.SN*) produce similar hazard curves, with the *Model2.SN* curve slightly higher for all exceedance rates. Hazard curves from these two models are much steeper compared to the other two models that assume normal probability distribution of $\ln(D)$ for displacement greater than about 1 m, resulting in progressively smaller and therefore more realistic prediction of displacement at increasingly low exceedance rates. Among the three new models, *Model1.NO* produces shallower hazard curves and much larger displacement at low exceedance rates, consistent with the discussion in Chapter 3 that models that assume normal probability distribution have a biased upper tail that results in large upper quantile displacement predictions, which controls the hazard at low exceedance rate. As displacement approaches zero, hazard curves from all models converge to an exceedance rate (intercept on the vertical axis in Figure 4.1) that equals the product of the applied event rate (7.14×10^{-3}), the probability term $P_2[D \neq 0 | A_{site,i}, z, sr \neq 0]$ (assumed to be 1), the probability term $P_1[D \neq 0 | z, sr \neq 0]$ (assumed to be 1), and the probability term $P[sr \neq 0 | m]$ (≈ 0.865 from Equation (A.3) for $m = 7$).

4.2.2 M 5.8 scenario example in IAEA benchmarking study

The **M** 5.8 scenario in the IAEA benchmarking study (Valentini et al., 2021) is developed based on fault geometry and activity data of the 5-km long Suizenji fault, a small cross fault near the Fudagawa fault zone that produced the 2016 **M** 7.0 Kumamoto earthquake in Japan. For this scenario, it is assumed that the fault ruptures repetitively with a **M** 5.8 characteristic earthquake at a mean rate of 23.3×10^{-5} events per year, and the site is located on the fault at $l/L = 0.39$.

In Figure 4.2, hazard curves calculated using the preferred model, *Model4.nEMG*, the P11 bilinear model, and the two interim models are compared. Compared to the P11 bilinear model, the new models predict lower displacement at all exceedance rates. This is because the bilinear magnitude scaling in the new models has a stronger magnitude scaling for small magnitudes resulting in lower predicted displacement values for the **M** 5.8 scenario, consistent with patterns shown in Figures 3.7, 3.11, and 3.17. Hazard curves from the new models have shallower slopes than the P11 bilinear model because of the larger variance of the random intercept depicted in

Figures 3.6, 3.9, and 3.15. Again, *Model2.SN* and *Model4.nEMG* yield similar hazard curves, with the nEMG hazard curve slightly lower and the amount of difference increases slightly as exceedance rate decreases.

4.2.3 Hypothetical M 8.3 scenario

Hazard curves for a hypothetical M 8.3 scenario were calculated to evaluate the extrapolation of the new nEMG model to large magnitudes. This example assumes that the M 8.3 earthquake repeats once every 2500 years (annual rate of 4×10^{-4}), and the site is located at $I/L = 0.4055$ (same as the M 7 scenario discussed in Section 4.2.1).

For context, the largest magnitude in the USGS ShakeMap BSSC2014 Scenario Catalog is 8.2 (see <https://earthquake.usgs.gov/scenarios/catalog/bssc2014/>). The catalog was developed for the Building Seismic Safety Council to provide ground motions for the deterministic component of ground motion hazard analyses required in the seismic design codes. The catalog includes a representative subset of ruptures that are based on the same earthquake source models as the USGS 2014 update of the National Seismic Hazard Model (Petersen et al., 2014). The M 8.2 scenario earthquake is on the southern San Andreas fault. Based on the Uniform California Earthquake Rupture Forecast, Version 3 (Field et al., 2013), the longest observed average paleoseismic mean recurrence interval (MRI) of all southern San Andreas fault sections is about 277 yr, and the longest model predicted MRI is 306 yr. The average moment rate for all fault sections ranges from 9.89×10^{16} Nm/yr to 9.88×10^{17} Nm/yr, which corresponds to an MRI ranging from 2,266 yr to 22,636 yr, respectively, assuming all energy is released by the characteristic M 8.2 earthquakes. This is not a realistic assumption because we know the southern San Andreas fault has ruptured more frequently with smaller magnitude earthquakes. Nevertheless, it is important to discuss hazard curves for large magnitudes because a comprehensive hazard analysis would include a more realistic magnitude frequency relation that necessitates the analysis of very large magnitudes.

In Figure 4.3, hazard curves calculated using the preferred model, *Model4.nEMG*, the P11 bilinear model, and the two interim models are compared. Differences in hazard curves obtained using different FDMs are amplified at such large magnitude, although the *Model4.nEMG* and *Model2.SN* curves are still very similar. These two models render much smaller displacement than P11's bilinear model, which renders over 50 m of displacement at the exceedance rate of 10^{-5} . The overall larger displacement from the P11 model reflects the earlier observation that the linear magnitude scaling does not extrapolate well to large magnitudes compared to bilinear magnitude scaling. With improved magnitude scaling and statistical analysis that considers earthquake to earthquake variability, the *Model1.NO* curve is improved compared to the P11 curve. However, it still yields seemingly unreasonably large displacement (never observed historically or in paleoseismic trenches from a single event) at low exceedance rate due to its normal assumption. In contrast, for such a large magnitude, the skew normal and the nEMG models render much more

reasonable displacement hazard. The nEMG model estimate is lower than the skew-normal model for all exceedance rates.

4.2.4 Scenario examples with model epistemic uncertainty

Estimation uncertainty in magnitude scaling and asymmetric slip profiles are discussed in Sections 3.5.1 and 3.6.2, respectively. Hazard implications of these epistemic uncertainties are illustrated in this section using scenario examples.

Section 3.3.3 states that the baseline nEMG model assumes $m_3 = 7.1$ for bilinear magnitude scaling. Estimation uncertainty is assessed in Section 3.5.1, and the effect of assessed uncertainty on predicted displacement distribution is illustrated using three additional m_3 values: 7.32, 6.75, and 6.40. In Figures 4.4a through 4.4c, hazard curves using these alternative m_3 are compared for the three scenario examples discussed in Sections 4.2.1 through 4.2.3, respectively. Whereas alternative m_3 values lead to hazard curves that are quite similar for the **M** 7.0 scenario (Figure 4.4a), significant differences are observed for the **M** 5.8 and **M** 8.3 scenarios. In general, the choice of m_3 will make significant differences for magnitudes away from the center of the data range and when the model is extrapolated beyond the data range, particularly when used for magnitude greater than about 8.0. For these situations, it is important to incorporate this epistemic uncertainty in hazard analysis (see Section 3.5.1.2).

The effect of asymmetric slip profile shape is illustrated in Figure 4.5a through 4.5d for the **M** 7.0 scenario, assuming a site on the fault at l/L of 0.5, 0.4, 0.3, and 0.1, respectively. The three alternative slip profiles include a symmetric profile that peaks at the center of the fault (i.e., $l/L = 0.5$), an asymmetric profile that peaks to the left (with respect to the center), and an asymmetric profile that peaks to the right. These figures show that whether slip profile is assumed to be symmetric or asymmetric and whether an asymmetric profile peaks to the left or to the right of the center of fault trace have increasingly more substantial effect on estimated hazard curves as the site moves from the center of the fault trace toward the ends of the fault and as exceedance rate decreases.

4.2.5 Example in IAEA benchmarking study with source uncertainty

This section illustrates the application of the *Model4.nEMG* model for a benchmarking study case that involves a seismic source logic tree developed by the IAEA (Valentini et al., 2021) to account for epistemic uncertainty in earthquake magnitude and occurrence rate. Fault displacement hazard from the Futagawa fault zone at a fictitious site on the fault is calculated.

In the IAEA source model, the Futagawa fault zone consists of three segments as shown in Figure 4.6a: the Futagawa, Uto, and Uto-Hanto-North segments (labeled as ①, ②, and ③, respectively). The candidate area or site is moved to be on the fault instead of what is shown in the figure. Hazard from Takano-Shirahata segment is ignored. The site is near the junction of the Uto

and Futagawa segments. Seismic activity on the Futagawa fault zone is characterized by 4 rupture scenarios involving a single segment rupture and 3 different combinations of multi-segment ruptures as shown in Figure 4.6b. The central magnitude and l/L values for each scenario are also shown in this figure.

The central magnitude and the mean rate values for each scenario are indicated in Figure 4.6c. Uncertainty in these values is implemented via the logic tree nodes and branches illustrated in the same figure. There are 3 magnitude branches and 3 rate branches, resulting in 9 end nodes for each scenario. Branch magnitude and rate values and weights for the Uto scenario are given as an example. For all scenarios, including the Uto scenario, magnitude uncertainty reflects a factor of 2 uncertainty in seismic moment, corresponding to an increase or decrease by 0.2 magnitude unit from the central values. Mean rate uncertainty is a factor of 3 above or below the central values. Branch weights for other scenarios are similar to the Uto scenario. That is, for both magnitude and rate, the central branch is weighted 60% and the upper and lower branches are weighted 20% each, respectively. The logic trees for the four scenarios are additive and not alternative (i.e., branch weights for each scenario add to 1). The total hazard is the sum of hazards from the four scenarios.

Figure 4.7a shows the mean and percentiles of fault displacement hazard curves using only the *Model4.nEMG* model. Simulated mean and percentile hazard curves are calculated using Monte Carlo simulation. The Monte Carlo sampling includes two steps. In the first step, 10,000 hazard curves are sampled for each scenario from its 9 logic-tree end-node hazard curves. The number of times each hazard curve is sampled is determined by its logic tree weight. In the second step, 10,000 total hazard curves are obtained by randomly selecting one hazard curve from each of the 4 scenario samples and adding them up. Logic-tree weights need not be considered in the summation because hazard curves in the scenario samples are already proportioned based on logic-tree weights in the first step. Finally, the 10,000 total hazard curves are ranked and percentile values determined at each displacement level.

The simulated mean recovers the calculated mean well as seen in Figure 4.7a, implying adequate sampling. Uncertainty range is quite large. The 5th to 95th percentile hazard ranges from 7×10^{-5} to 4×10^{-4} at the 10-cm displacement level and from about 1.8×10^{-5} to 8×10^{-4} at the 1-m displacement level. The mean displacement is approximately 0.35 m, 2.5 m, and 6.0 m at exceedance rates of 10^{-4} , 10^{-5} , and 10^{-6} respectively. The uncertainty ranges in these hazard curves reflect alternative magnitude and event rate values given by the logic tree shown in Figure 4.6c, which only covers part of the sources of epistemic uncertainty. One would expect an even large range if other uncertainties were considered, for example uncertainties in FDMs (i.e., using a range of models), in magnitude-frequency distribution, and in the estimation of surface rupture probability.

The mean hazard curve using the P11 bilinear model is shown for comparison (labeled P11 Bilinear) in Figure 4.7a. The P11 model produces higher total mean hazard compared to the

Model4.nEMG model, particularly at low annual exceedance rates mainly because of its normal probability assumption with larger standard deviation.

Figure 4.7b plots mean hazard curves from each of the four scenarios and the total mean hazard curves. The small and more frequent Uto scenario contributes the most to the total hazard at hazard level higher than about 10^{-5} , below which the controlling scenario is the **M** 7.2 scenario that involves all three segments of the Futagawa fault zone rupturing together.

For comparison, total and scenario hazard curves from P11 bilinear model are also plotted on Figure 4.7b. P11 hazard curves are higher than the nEMG curve for Uto **M** 6.5 scenario, but lower than the corresponding nEMG curves for the other three larger magnitude scenarios, resulting in coincidentally comparable total hazards from these two models at small exceedance rates. At larger exceedance rate, the P11 model produces overall higher total hazards.

Note in Chapter 3 we recommend adding epistemic uncertainty in m_1 when the dominant source is **M** < 6.5. This is not done in examples shown in this chapter to keep these examples simple and to make comparisons of hazard results from different FEMs easier.

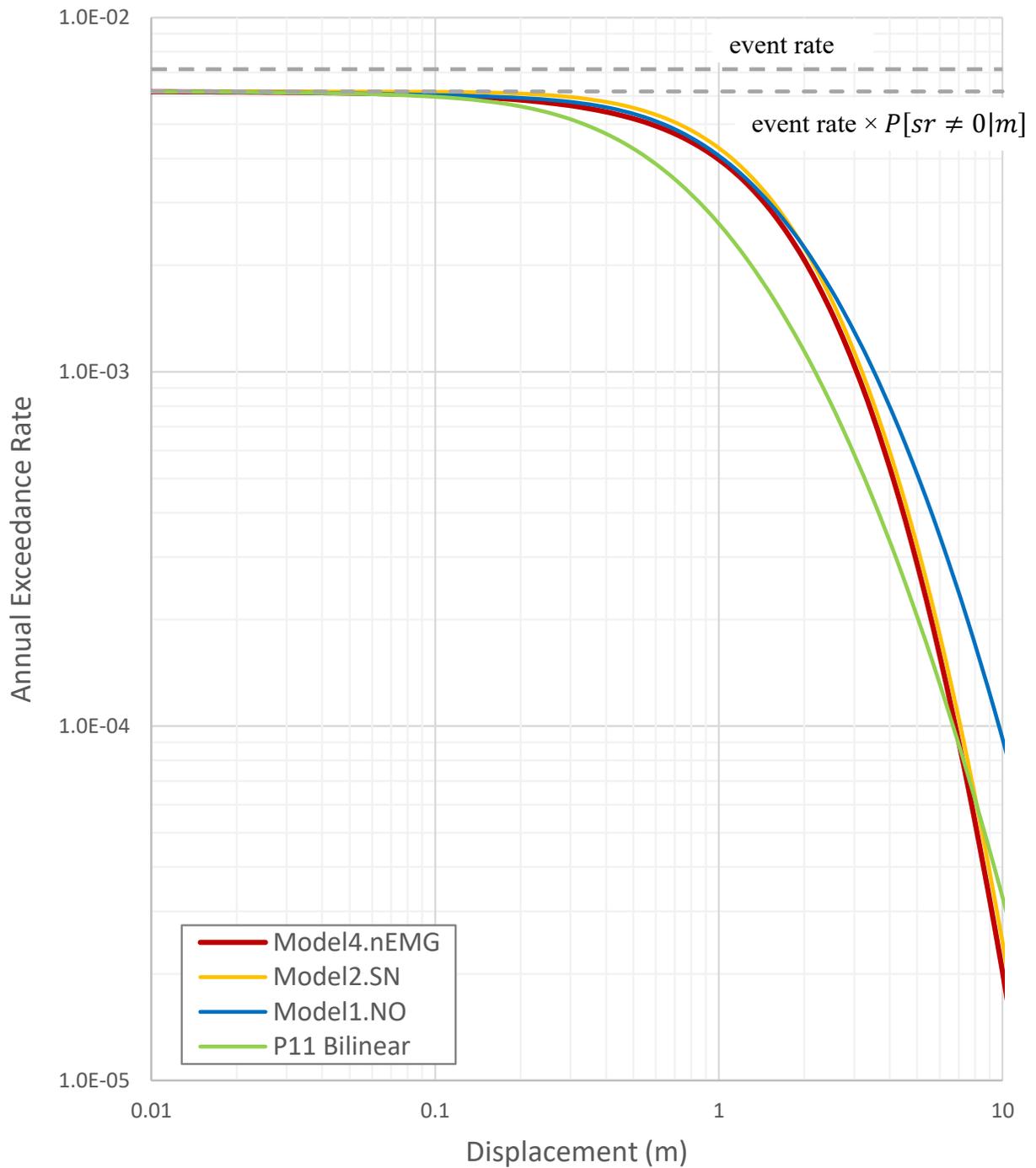


Figure 4.1. Comparison of hazard curves for the magnitude (**M**) 7.0 scenario earthquake using the preferred negative exponentially modified Gaussian (nEMG) model (*Model4.nEMG*), the Petersen et al. (2011) bilinear model (P11 Bilinear), and two interim models (*Model1.NO* and *Model2.SN*). $P[sr \neq 0|m]$ is the probability of having surface rupture (i.e., $sr \neq 0$) given that a magnitude m earthquake occurs.

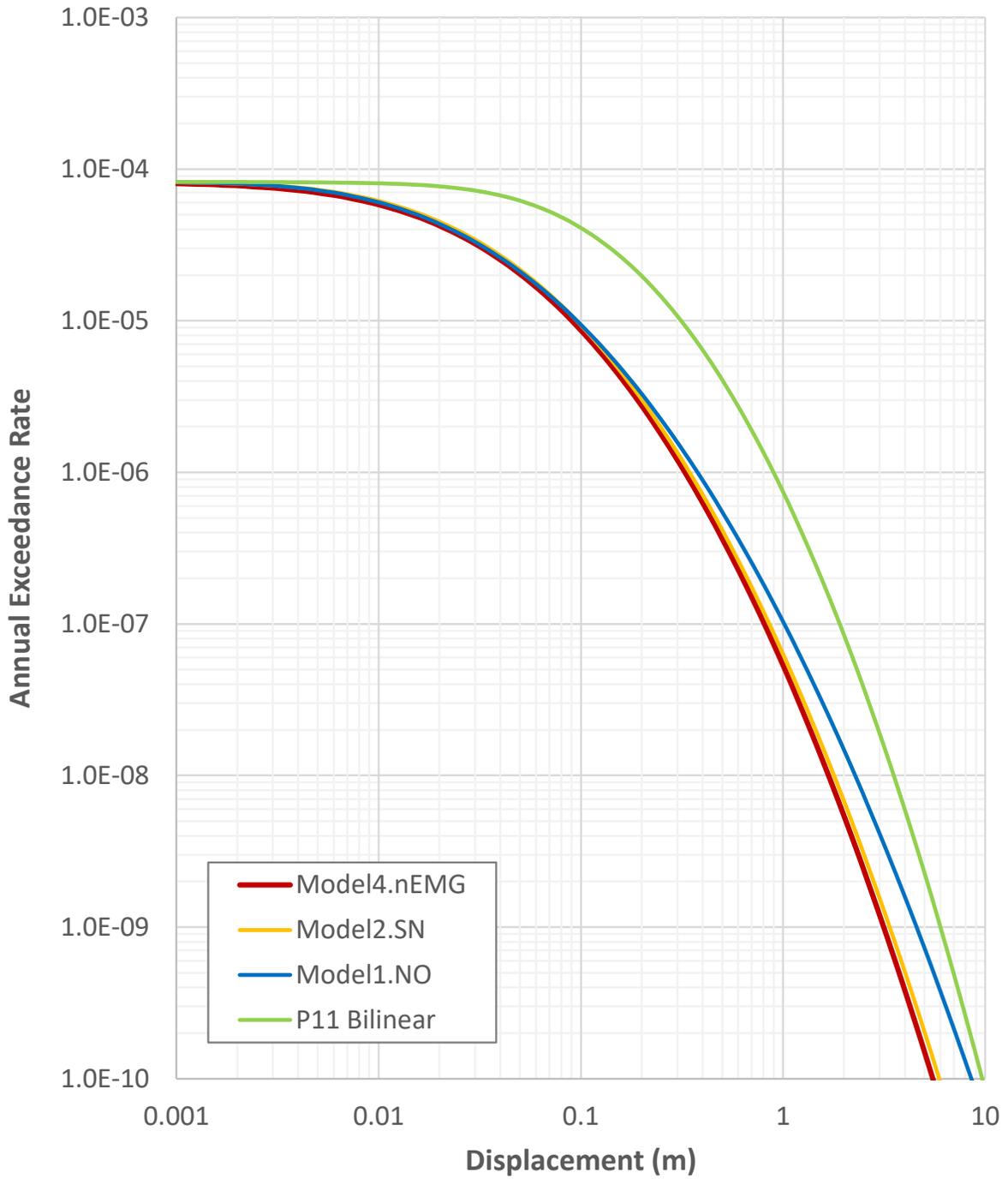


Figure 4.2. Comparison of hazard curves for the **M** 5.8 scenario earthquake using the preferred negative exponentially modified Gaussian (nEMG) model, the Petersen et al. (2011) bilinear model (P11 Bilinear), and two interim models (*Model1.NO* and *Model2.SN*).

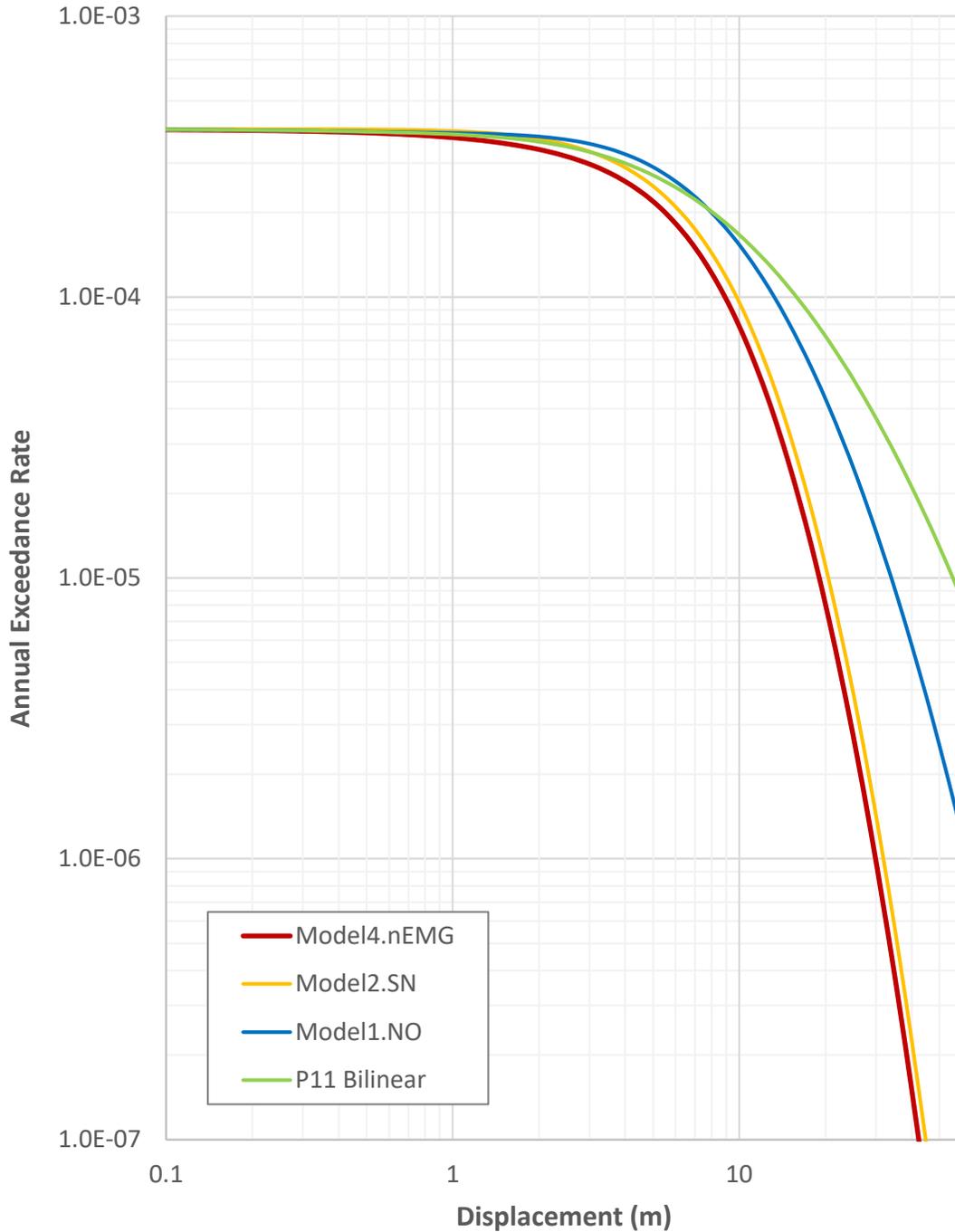


Figure 4.3. Comparison of hazard curves for the hypothetical magnitude (**M**) 8.3 scenario using the preferred negative exponentially modified Gaussian (nEMG) model, the Petersen et al. (2011) bilinear model (P11 Bilinear), and two interim models (*Model1.NO* and *Model2.SN*).

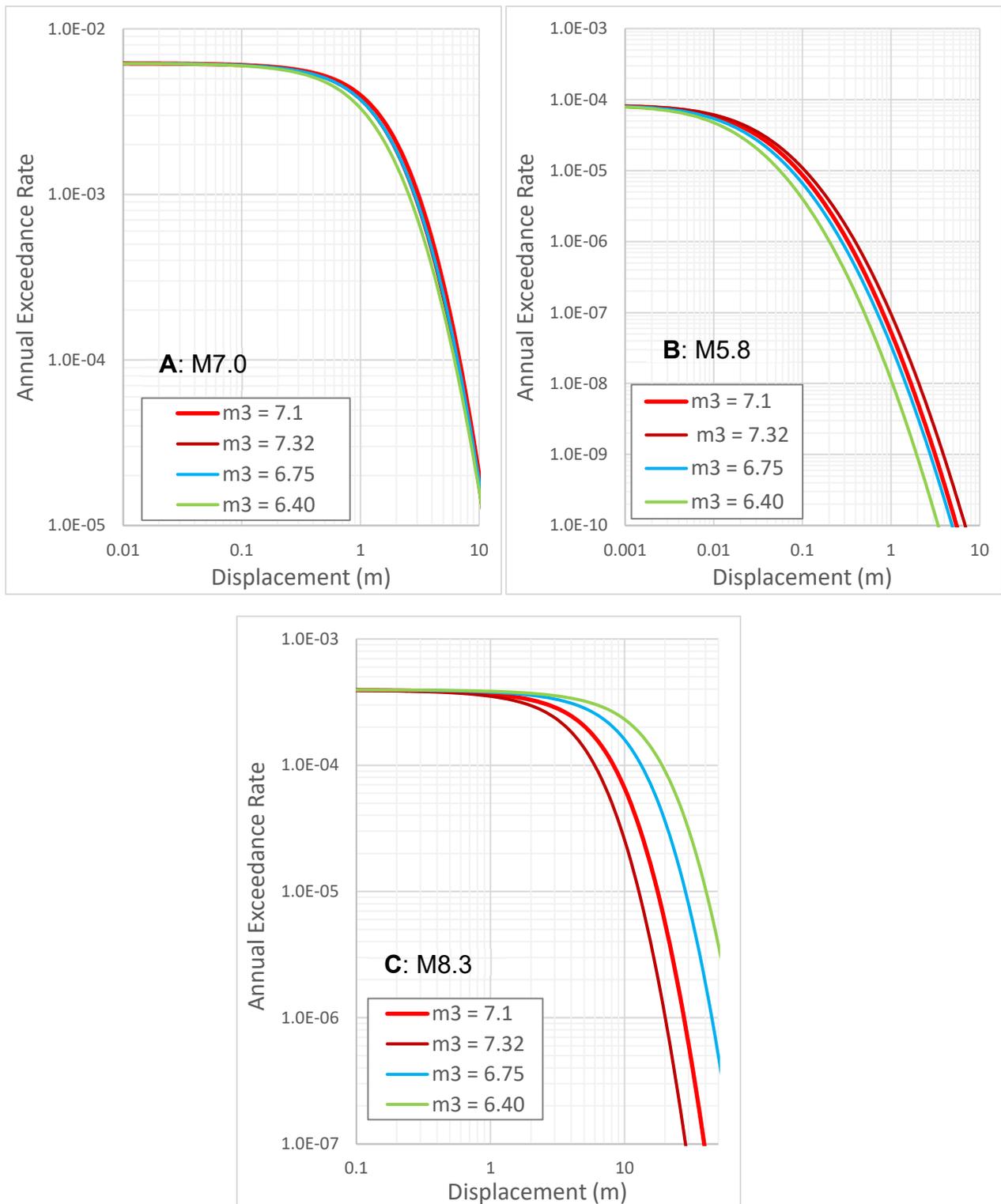


Figure 4.4. Hazard curves showing effects of alternative m_3 values in models using negative exponentially modified Gaussian (nEMG) distribution. (A) Magnitude (**M**) 7.0 scenario in Petersen et al. (2011). (B) **M** 5.8 scenario example in International Atomic Energy Agency (IAEA) benchmarking study (Valentini et al., 2021). (C) Hypothetical **M** 8.3 scenario.

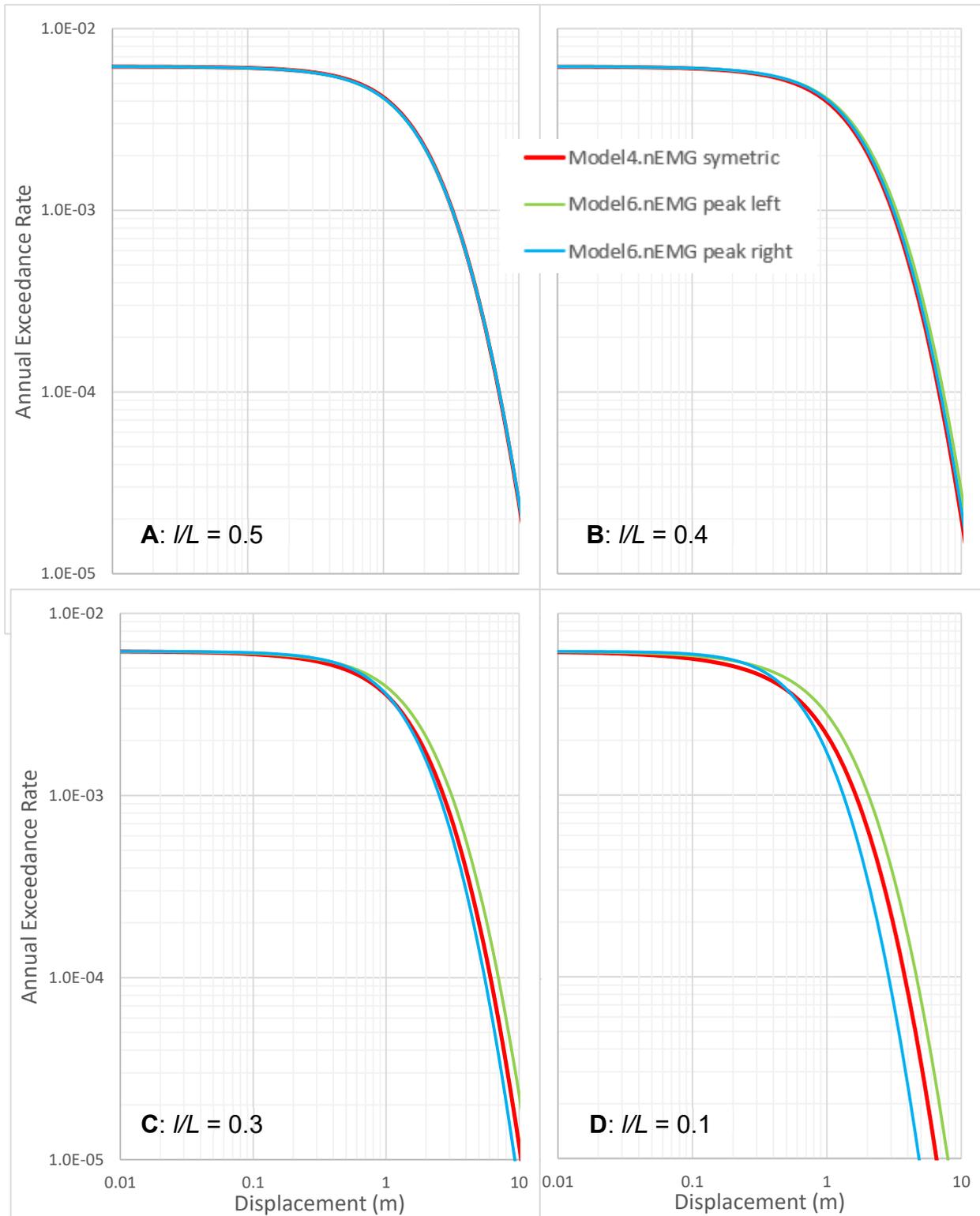


Figure 4.5. Hazard curves showing effects of alternative assumptions for along-strike slip distribution using the magnitude (M) 7.0 scenario as an example. (A) $I/L = 0.5$. (B) $I/L = 0.4$. (C) $I/L = 0.3$. (D) $I/L = 0.1$. Variable I is the along-main-trace position and L is the length of the main trace (see Figure 2.1).

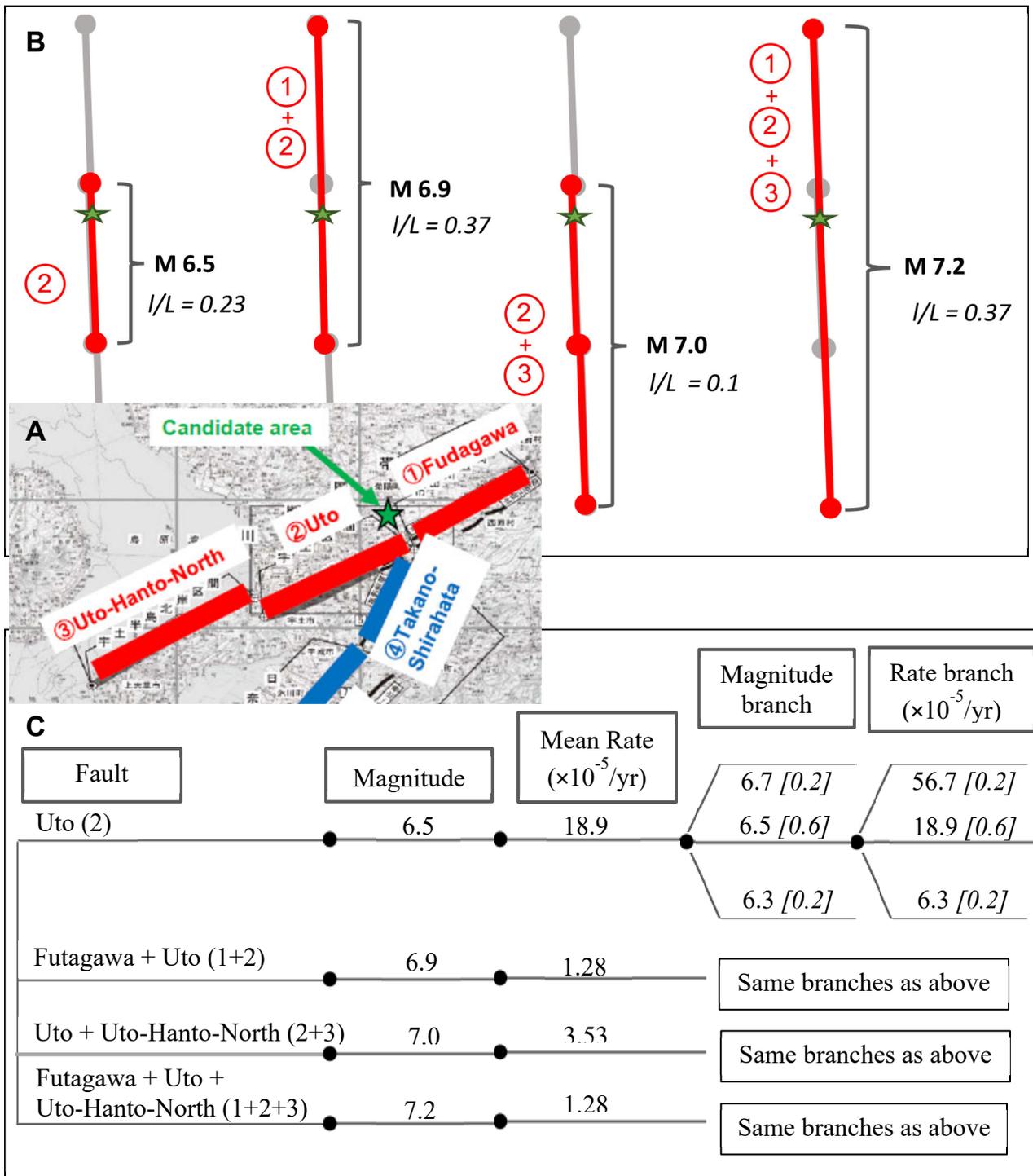


Figure 4.6. Example in the International Atomic Energy Agency (IAEA) benchmarking study with source model logic tree (modified from Valentini et al., 2021). (A) Three segments of the Fudagawa fault zone. (B) Rupture scenarios. (C) Source logic tree with epistemic uncertainty in magnitude and event rate. M is earthquake magnitude; I is the along-main-trace position, and L is the length of the main trace (see Figure 2.1).

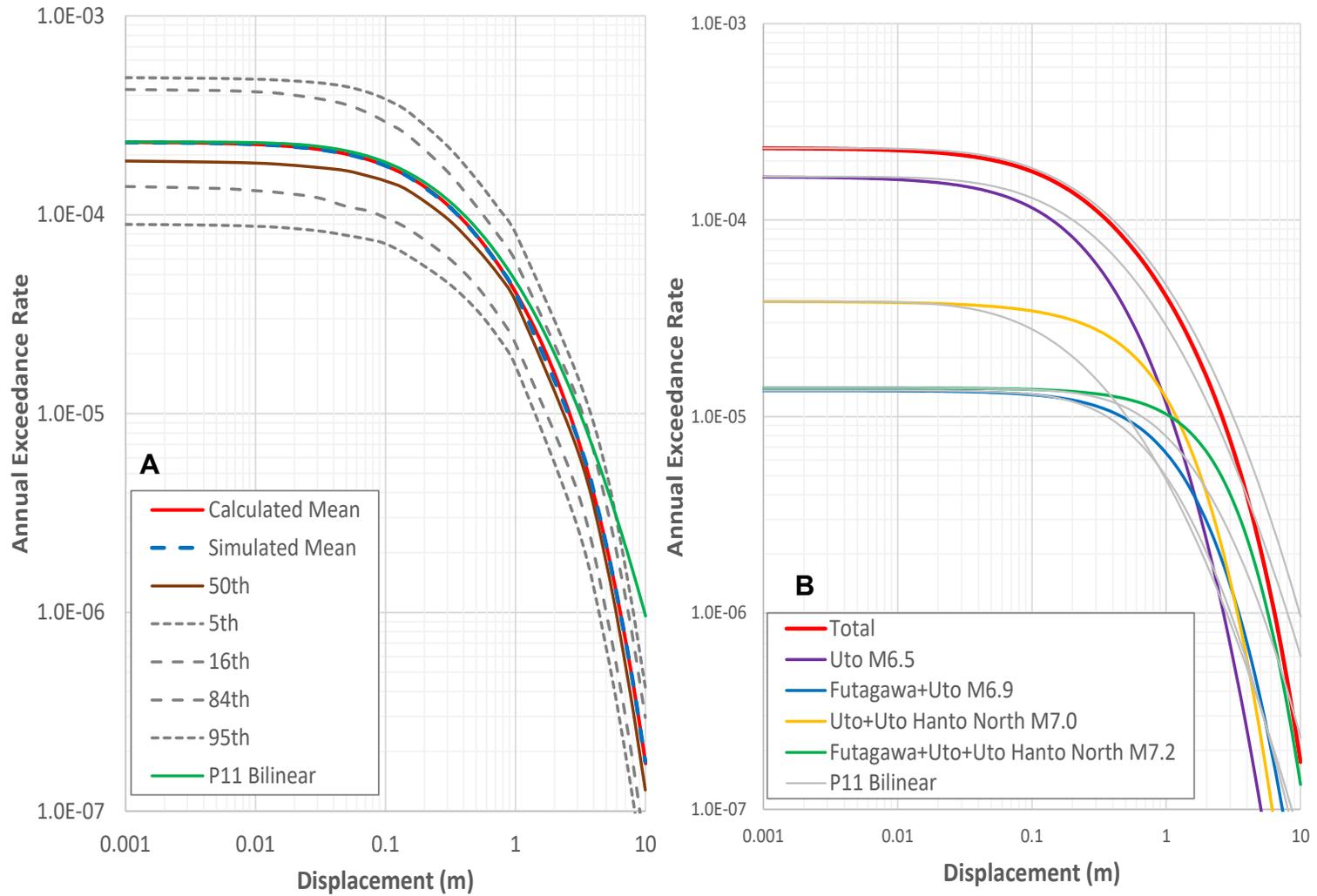


Figure 4.7. Comparison of Petersen et al. (2011; P11) bilinear and negative exponentially modified Gaussian (nEMG) hazard curves for a site on the Fudagawa fault zone in Japan. Calculated mean hazard curve is the sum of weighted mean of all scenario curves. (A) Mean and percentile hazard curves. (B) Scenario and total hazard curves. M is earthquake magnitude.

5 Summary and Future Work

5.1 SUMMARY

This report documents the development of several fault displacement models (FDMs) for strike-slip surface rupture as results of our effort to update the FDM in Petersen et al. (2011, P11). These models predict the probabilistic distribution of aggregated principal net displacement across a main surface rupture trace. The response variable in these FDMs is the natural logarithm of the aggregated principal displacement, $\ln(D)$. Predictor variables include moment magnitude (M) and normalized location (l/L or $l2L$) along the main rupture trace, which is developed from mapped rupture traces in ArcGIS using the least-cost path (LCP) analysis approach. As is done in P11 and in Chen and Petersen (2019), the aggregated displacement can be partitioned to subparallel fault traces in hazard application.

A simple update of P11 (*Model0.NO*), three interim updates (*Model1.NO*, *Model2.SN*, and *Model3.ST*) using different assumed probability distributions for $\ln(D)$, and the preferred model (*Model4.nEMG*) are presented in Chapter 3. Through residual diagnostics of the simple update, we identify several crucial refinements to P11's formulations, and these refinements are incorporated into the three interim updates and the preferred model. *Model1.NO* includes these refinements, while maintaining P11's normal probability distribution assumption. Efforts to handle the observed non-normal distributions of model residuals lead to the adoption of skew normal and skew- t distributions in *Model2.SN* and *Model3.ST*, respectively. These two non-normal distributions have increasingly larger flexibility in regulating the shape of their density functions to improve fits to non-normal distribution of slip data, at the cost of less interpretability of distribution parameters and higher complexity in forward application. Our preferred model (*Model4.nEMG*) largely avoids such sacrifices. The interim models and the preferred model predict similar quantile displacements, and their upper (95th) percentile displacement predictions are more physically reasonable compared to those by P11 and by the simple update of P11 (which is without the benefits of model reformulations). Furthermore, the 95th percentile displacements predicted by our models are commensurate with the observed maximum displacement from several large historically observed surface rupture earthquakes. We judge that the interim updates are valid

models and can be used as alternatives in PFDHA to account for the epistemic uncertainty in the choice of probability distribution, if needed.

We investigate epistemic uncertainty in the **M**-scaling with displacement and provide three alternatives of *Model4.nEMG* conditional on three different m_3 values: 6.4, 6.75, and 7.32, in addition to the $m_3 = 7.1$ in *Model4.nEMG*. These alternative FDMs (*Model5.1.nEMG*, *Model5.2.nEMG*, and *Model5.3.nEMG*) can be incorporated via a logic-tree approach to model the uncertainty in **M**-scaling in hazard analysis. When the dominant rupture source is of magnitude (**M**) < 6.5, due to insufficient earthquakes in such magnitude range in the FDHI database to constrain the **M**-scaling, we also recommend inclusion of an additional uncertainty in the **M**-scaling rate of lower magnitudes (the coefficient m_1) to reflect the statistical estimation uncertainty of m_1 .

Like the P11 model, the interim models and *Model4.nEMG* assume a symmetric along-strike slip profile about the rupture middle point ($l/L = 0.5$). *Model6.nEMG* extends the preferred model to account for asymmetric slip profile in future rupture. In PFDHA, a hazard analyst can choose to use *Model4.nEMG*, *Model6.nEMG* (if the along-strike profile skew direction is known), or both models and skew directions via a logic tree to account for uncertainty in the skew-direction of future slip profile.

Comparisons of FDMs are given in Section 3.4, including reasons for preferring *Model4.nEMG* over other non-normal models. To reiterate, the reasons are as follows: (1) intuitive connection to the rupture mechanism responsible for the left-skewed distribution of $\ln(D)$ data, (2) simple interpretation of distribution parameters, (3) better fit to $\ln(D)$ data compared to interim models, and (4) analytic expressions of CDF and PDF, for both conditional and compound nEMG distributions, allow for a relatively straight-forward implementation in engineering applications.

The applicable magnitude range of our models is from **M** 6 to **M** 7.9, based solely on the magnitude range of data used in our models. However, analysis of $\ln(D_{ave})$ (treated as a surrogate of $\ln(D)$) from non-FDHI data sources indicates that the **M**-scaling of our models adequately captures $\ln(D_{ave})$ in the range of $7.9 < \mathbf{M} \leq 8.3$. This can potentially serve as a basis for extending the upper bound of applicable range to **M** 8.3 in hazard analysis.

Hazard applications of our models are demonstrated in Chapter 4 in the PFDHA framework established in P11. Hazard curves based on P11, interim, and the preferred models are compared and utilized to further illustrate the effects of FDM formulation, assumed probability distribution, epistemic uncertainty in **M**-scaling relation, and other assumptions adopted in our models.

5.2 POTENTIAL MODEL IMPROVEMENTS

Throughout Chapter 3, we mentioned potential future improvements to *Model4.nEMG* and topics that warrant further investigation. These include the following:

- Further investigation in the random effects on the scale and shape parameters of assumed probability distribution, which may provide insights into potential relations between rupture complexity and the observed level of variability in slip profile (section 3.2.3.2).
- Determination of the left truncation point of m_3 so that its truncated normal distribution can be used in developing logic tree branches and associated weighting schemes in hazard analyses to account for the epistemic uncertainty of predicted distribution of displacement due to the epistemic uncertainty in estimating m_3 (Section 3.5.1.2).
- Further development of along-strike asymmetric FDMs (*Model6.nEMG*), including along-strike asymmetry of ν parameter, epistemic uncertainty in \mathbf{M} -scaling, and implementation in PFDHA (Section 3.6.2).
- Resolution of discrepancy in m_1 in existing average displacement (D_{ave}) datasets when additional earthquakes in the $\mathbf{M} < 6.5$ range become available, particularly if consistent calculation of D_{ave} can be achieved (Section 3.6.5).
- Further investigation into spatial correlation in displacement data, in particular, the possible relation between observed spatial correlation (that is, $\ln(D)$ at two nearby points on the main trace are statistically correlated) and interior slip tapering, proper type of model residual for spatial correlation analysis, and statistical method for characterizing and modeling of spatial correlation in the context of nEMG distribution, and the possible earthquake-to-earthquake variation in the spatial correlation within individual earthquakes (Section 3.6.6).

5.3 FUTURE MODEL DEVELOPMENT

In addition to the FDMs for principal displacement, other component models in Equation (A.1) of Appendix A can be developed or improved using the FDHI database. Future model development work may include the following:

1. FDMs for the principal displacement for normal and reverse faults – The same approaches documented in this report can be used, including approaches for data selection, data processing to define reference coordinates, data processing to aggregate principal displacement across multiple subparallel rupture traces, and FDM development via regression of GAMLSS (see Section 3.2.3.2). Data from normal, normal-oblique, reverse, and reverse-oblique earthquakes will likely be analyzed separately initially. These data may be combined for statistical efficiency if initial analyses indicate that such combination is reasonable for the inference of parts of FDM that share commonality between different styles of faulting.
2. FDMs for distributed displacement – Data selection and calculation of coordinates relative to the LCP (i.e., u_{LCP} and t_{LCP}) have already been completed for distributed displacement for strike-slip earthquakes. We plan to update the P11 distributed

displacement model for strike slip faults. FDMs for distributed displacement for other styles of faulting can also be developed.

3. Surface rupture probability for principal displacement – This is the probability term proposed in P11 to account for gaps between principal rupture segments along the main rupture trace. Data needed to refine this term can be developed using the FDHI database.
4. Surface rupture probability for distributed displacement – This probability term is very important in assessing hazards from distributed ruptures (see P11 and Youngs et al., 2003). The cell counting approach used in Youngs et al. (2003) and in P11 can be followed. The slicing method used in Nurminen et al. (2020) and other potentially viable approaches can be examined and followed if applicable.

REFERENCES

- Abrahamson N. A., Youngs R. R. (1992). A stable algorithm for regression analyses using the random effects model. *Bulletin of the Seismological Society of America*, **82**(1), 505–510, doi: 10.1785/BSSA0820010505.
- AMEC Geomatrix (2010). *Geotechnical and Engineering Geological Investigation: California Memorial Stadium- Phase 1 Seismic and Program Improvements*, Project 0107660070.000005.
- American Nuclear Society (2015). *Criteria for Assessing Tectonic Surface Rupture and Deformation at Nuclear Facilities*, ANSI/ANS-2.30-2015, 50 pp.
- Anderson J. G., Biasi G. P., Angster S., Wesnousky S. G. (2021). Improved scaling relationships for seismic moment and average slip of strike-slip earthquakes incorporating fault-slip rate, fault width, and stress drop. *Bulletin of the Seismological Society of America*, **111**(5), 2379–2392, doi: 10.1785/0120210113
- Arellano-Valle R. B., Bolfarine H., Lachos V. H. (2005). Skew-normal linear mixed models. *Journal of Data Science*, **3**(4), 415–438, doi: 10.6339/JDS.2005.03(4).238.
- Azzalini A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**(2), 171–178.
- Azzalini A., Capitanio A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t*-distribution. *Journal of the Royal Statistical Society: Series B*, **65**(2), 367–389, doi: 10.1111/1467-9868.00391.
- Azzalini A., Capitanio A. (2014). *The Skew-Normal and Related Families*, Institute of Mathematical Statistics Monographs, Cambridge University Press: New York, 262 pp., doi: 10.1017/CBO9781139248891.
- Baize S., Nurminen F., Sarmiento A., Dawson T., Takao M., Scotti O., Azuma T., Boncio P., Champenois J., Cinti F. R., Civico R., Costa C., Guerrieri L., Marti E., McCalpin J., Okumura K., Villamor P. (2020). A worldwide and unified database of surface ruptures (SURE) for fault displacement hazard analysis. *Seismological Research Letters*, **91**(1), 499–520, doi: 10.1785/0220190144.
- Bozorgnia Y., Abrahamson N., Arcos B., Baize S., Boncio P., Chao S.-H., Chen R., Chiou B., Dawson T., Donahue J., Goulet C., Hanson K., Kottke A., Kuehn N., Kuo C.-H., Lavrentiadis G., Madugo D., Mazzoni S., Milliner C., Moss R., Nurminen F., Pace B., Petersen M., Sarmiento A., Shen A., Thomas K., Thompson S., Visini F., Wang Y., Youngs R. R. (2021). An overview of the fault displacement hazard initiative research program. *Seismological Research Letters*, **92**(2B), 1360–1361.
- Briney A. (2014). Overview of least cost path analysis. *Spatial Analysis GIS Lounge*, April 7, 2014. <https://www.gislounge.com/overview-least-cost-path-analysis/>. Accessed on March 4, 2021.
- Chang K. T. (2012). *Introduction to Geographic Information Systems*, 6th Edition. McGraw-Hill: New York, 248 pp.
- Chen R., Petersen M. D. (2011). Mapping fault displacement hazards for the ShakeOut scenario using semi-probabilistic and probabilistic approaches. *Earthquake Spectra*, **27**(2), 293–313, doi: 10.1193/1.3574226.
- Chen R., Petersen M. D. (2019). Improved implementation of rupture location uncertainty in fault displacement hazard assessment. *Bulletin of the Seismological Society of America*, **109**(5), 2132–2137, doi: 10.1785/0120180305.

- Chen R., Dawson T. E., Wills C. J. (2013). *Quantifying Surface Fault Rupture Location Uncertainties for Lifeline Crossings*, Final Technical Report, U.S. Geological Survey National Earthquake Hazards Reduction Program, Award Number G11AP20040, 51 pp.
- Chiou B. S. J., Youngs R. R. (2008). An NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra*, **24**(1), 173–215, doi: 10.1193/1.2894832.
- Chiou B. S. J., Youngs R. R. (2014). Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra*, **30**(3), 1117–1153, doi: 10.1193/072813EQS219M.
- Dunn P. L., Smyth G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**(3), 236–244, doi: 10.2307/1390802.
- DuRoss C. B., Gold R. D., Dawson T. E., Scharer K. M., Kendrick K.J., Akciz S.O., Angster S. J., Bachhuber J., Bacon S., Bennett S.E.K., et al. (2020). Surface displacement distributions for the July 2019 Ridgecrest, California, earthquake ruptures. *Bulletin of the Seismological Society of America*, **110**(4), 1400–1418, doi: 10.1785/0120200058.
- Field E. H., Biasi G. P., Bird P., Dawson T. E., Felzer K. R., Jackson D. D., Johnson K. M., Jordan T. H., Madden C., Michael A. J., Milner K. R., Page M. T., Parsons T., Powers P. M., Shaw B. E., Thatcher W. R., Weldon R.J. II, Zeng, Y. (2013). *Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3)—The Time-Independent Model*, U.S. Geological Survey Open-File Report 2013–1165, California Geological Survey Special Report 228, and Southern California Earthquake Center Publication 1792, 97 pp., doi: 10.3133/ofr20131165.
- Gori J., Rioul O. (2019). Regression to a linear lower bound with outliers: An exponentially modified Gaussian noise model. *27th European Signal Processing Conference (EUSIPCO)*, A Coruna, Spain, 5 pp, doi: 10.23919/EUSIPCO.2019.8902946.
- Haeussler, P. J. (2004). Surface rupture and slip distribution of the Denali and Totschunda faults in the 3 November 2002 M 7.9 earthquake, Alaska. *Bulletin of the Seismological Society of America*, **94**(6B), S23–S52, doi: 10.1785/0120040626.
- Hastie T. J., Tibshirani R. J. (1990). *Generalized additive models*, Chapman & Hall: London, 352 pp.
- International Atomic Energy Agency (IAEA) (2016). *Site Evaluation for Nuclear Installations – Safety Requirements*, Safety Standards Series No. NS-R-3 (Rev. 1), Vienna, Austria, 27 pp.
- International Atomic Energy Agency (IAEA) (2021). *An Introduction to Probabilistic Fault Displacement Hazard Analysis in Site Evaluation for Existing Nuclear Installations*, IAEA-TECDOC-1987, Vienna, Austria, 144 pp.
- Japan Nuclear Safety Institute (2013). *Assessment Methods for Nuclear Power Plant Against Fault Displacement* (Provisional Translation of Main Text), JANSI-FDE-03, September.
- Kanninen M. F., Popelar C. H. (1985). *Advanced Fracture Mechanics*, Oxford University Press: Oxford.
- Kneib T., Silbersdorff A., Säfken B. (2021). Rage against the mean – A review of distributional regression approaches. *Econometrics and Statistics*, **26**, 99–123, doi: 10.1016/j.ecosta.2021.07.006.
- Koenker R. (2005). *Quantile Regression*, Cambridge University Press: Cambridge, 349 pp., doi: 10.1017/CBO9780511754098.
- Kuehn N., Kottke A., Madugo C., Sarmiento A., Bozorgnia Y. (2022). *UCLA-PG&E Fault Displacement Model*, Report GIRS-2022-06, University of California, Los Angeles, 186 pp., doi: 10.34948/N3X59H.
- Kumar S., Wesnousky S. G., Rockwell T. K., Briggs R. W., Thakur V. C., Jayangondaperumal R. (2006). Paleoseismic evidence of great surface rupture earthquakes along the Indian Himalaya. *Journal of Geophysical Research: Solid Earth*, **111**(B3), B03304, doi: 10.1029/2004JB003309.

- Lovison G., Schindler C. (2014). Separate regression modelling of the Gaussian and Exponential components of an EMG response from respiratory physiology. *Proceedings of the 29th International Workshop on Statistical Modeling*, 189–194.
- Manighetti I., Campillo M., Sammis C., Mai P. M., King G. (2005). Evidence for self-similar, triangular slip distributions on earthquakes: Implications for earthquake and fault mechanics. *Journal of Geophysical Research: Solid Earth*, **110**(B5), B05302, doi: 10.1029/2004JB003174.
- Manighetti I., Caulet C., De Barros L., Perrin C., Cappa F., Gaudemer Y. (2015). Generic along-strike segmentation of Afar normal faults, East Africa: Implications on fault growth and stress heterogeneity on seismogenic fault planes. *Geochemistry, Geophysics, Geosystems*, **16**(2), 443–467. doi: 10.1002/2014GC005691.
- Martel S. J., Shacat C. (2006). Mechanics and interpretations of fault slip. In: R. Abercrombie, A. McGarr, H. Kanamori, G. Di Toro (eds.). *Earthquakes: Radiated Energy and the Physics of Faulting*, Geophysical Monograph Series, Volume 170, American Geophysical Union: Washington, D.C., pp. 207–215, doi: 10.1029/170GM21.
- Martin P.M., Thingbaijam K.K.S., (2014). SRCMOD: An online database of finite-fault rupture models. *Seismological Research Letters*, **85**(6), 1348–1357, doi: 10.1785/0220140077.
- McCullagh P., Nelder J. A. (1989). *Generalized Linear Models*, Second Edition, Monographs on Statistics and Applied Probability, Vol. 37. Chapman & Hall/CRC: Boca Raton, 532 pp., doi: 10.1201/9780203753736.
- Milliner C., Avouac J. P., Chen R., Aati S., Chiou B., Donnellan A., Dawson T., Madugo C., Dolan J. F. (2020). Development of a geodetic-based probabilistic fault displacement hazard analysis using near-field geodetic imaging data. Abstract T042–07 presented at 2020 AGU Fall Meeting, 1-17 Dec.
- Milliner C. W., Dolan J. F., Hollingsworth J., Leprince S., Ayoub F., Sammis C. G. (2016). Comparison of coseismic near-field and off-fault surface deformation patterns of the 1992 Mw 7.3 Landers and 1999 Mw 7.1 Hector Mine earthquakes: Implications for controls on the distribution of surface strain. *Geophysical Research Letters*, **43**(19), 10115–10124, doi: 10.1002/2016GL069841.
- Milliner C. W., Dolan J. F., Hollingsworth J., Leprince S., Ayoub F., Sammis C. G. (2015). Quantifying near-field and off-fault deformation patterns of the 1992 Mw 7.3 Landers earthquake. *Geochemistry, Geophysics, Geosystems*, **16**(5), 1577–1598 doi: 10.1002/2014GC005693.
- Moss R. E. S., Ross Z. E. (2011). Probabilistic fault displacement hazard analysis for reverse faults. *Bulletin of the Seismological Society of America*, **101**(4), 1542–1553, doi: 10.1785/0120100248.
- Moss R., Thompson S., Kuo C. H., Younesi K., Chao S. H. (2022). *Reverse Fault PFDHA*, Report GIRS-2022-05, University of California, Los Angeles, 124 pp., doi: 10.34948/N3F595.
- Nurminen F., Boncio P., Visini F., Pace B., Valentini A., Baize S., Scotti O. (2020). Probability of occurrence and displacement regression of distributed surface rupturing for reverse earthquakes, *Frontiers in Earth Science*, **8**, 581605, doi: 10.3389/feart.2020.581605.
- Petersen M. D., Dawson T. E., Chen R., Cao T., Wills C. J., Schwartz D. P., Frankel A. D. (2011). Fault displacement hazard for strike-slip faults. *Bulletin of the Seismological Society of America*, **101**(2), 805–825, doi: 10.1785/0120100035.
- Petersen M. D., Moschetti M. P., Powers P. M., Mueller C. S., Haller K. M., Frankel A. D., Zeng Y., Rezaeian S., Harmsen S. C., Boyd O. S., Field N., Chen R., Rukstales K. S., Luco N., Wheeler R. L., Williams R. A., 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., Shumway A. M., Powers P. M., Mueller C. S., Moschetti M. P., Frankel A. D., Rezaeian S., McNamara D. E., Luco N., Boyd O. S., Rukstales K. S., Jaiswal K. S., Thompson E. M., Hoover S. M.,

- Clayton B. S., Field E. H., 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.
- Pinheiro J. C, Bates D. M. (2000). *Mixed-Effects Models in S and S-PLUS*, Springer: New York, 528 pp., doi: 10.1007/b98882.
- Rigby R. A., Stasinopoulos M. D., Heller G. Z., Bastiani F. De (2020). *Distribution for Modelling Location, Scale, and Shape: Using GAMLSS in R*, Chapman & Hall/CRC: New York, 588 pp., doi: 10.1201/9780429298547.
- Rizzo P. C. (2013a). *Technical Report, Probabilistic Fault Displacement Hazard Analysis, Krško East and West Sites, Proposed Krško 2 Nuclear Power Plant, Krško, Slovenia, Revision 1*, Paul C. Rizzo Associates, Inc., Pittsburgh, Pennsylvania, USA.
- Rizzo P. C. (2013b). *Final Technical Report, Sensitivity analysis, Probabilistic Fault Displacement Hazard Analysis, Krško East and West sites, Proposed Krško 2 Nuclear Power Plant, Krško, Slovenia, Revision 1*, Paul C. Rizzo Associates, Inc., Pittsburgh, Pennsylvania, USA.
- Rockwell T. K., Klinger Y. (2013) Surface rupture and slip distribution of the 1940 Imperial Valley earthquake, Imperial fault, southern California: Implications for rupture segmentation and dynamics. *Bulletin of the Seismological Society of America*, **103**(2A), 629–640, doi: 10.1785/0120120192.
- Rodgers D. W., Little, T. A. (2006). World's largest coseismic strike-slip offset: The 1855 rupture of the Wairarapa Fault, New Zealand, and implications for displacement/length scaling of continental earthquakes. *Journal of Geophysical Research: Solid Earth*, **111**(B12), B12408, doi: [10.1029/2005JB004065](https://doi.org/10.1029/2005JB004065).
- Sarmiento A., Madugo D., Bozorgnia Y., Shen A., Mazzoni S., Lavrentiadis G., Dawson T., Madugo C., Kottke A., Thompson S., Baize S., Milliner C., Nurminen F., Boncio P., Visini F. (2021). *Fault Displacement Hazard Initiative Database*, UCLA B. John Garrick Institute for the Risk Sciences, Report GIRS-2021-08, Revision 3 dated 19 July 2022, 97 pp., doi: 10.34948/N36P48.
- Scholz, C. H. (2019). *The Mechanics of Earthquakes and Faulting*, 3rd Edition, Cambridge University Press: Cambridge, 519 pp., doi: 10.1017/9781316681473.
- Scholz, C. H., Lawler T. M. (2004). Slip tapers at the tips of faults and earthquake ruptures. *Geophysical Research Letters*, **31**(21), L21609, doi: [10.1029/2004GL021030](https://doi.org/10.1029/2004GL021030).
- Spudich P., Chiou B. (2015). *Strike-Parallel and Strike-Normal Coordinate System Around Geometrically Complicated Rupture Traces: Use by NGA-West2 and Further Improvements*, U.S. Geological Survey Open-File Report 2015-1028, 20 pp., doi: 10.3133/ofr20151028.
- Stasinopoulos M. D., Rigby R. A., Heller G. Z., Voudouris V., Bastiani F. D. (2017). *Flexible Regression and Smoothing: Using GAMLSS in R*, Chapman & Hall/CRC: New York, 571 pp., doi: 10.1201/b21973.
- Takao M., Tsuchiyama J., Annaka T., Kurita T. (2013). Application of probabilistic fault displacement hazard analysis in Japan. *Journal of the Japanese Association of Earthquake Engineering*, **13**(1), 17–36, doi: 10.5610/jaee.13.17.
- U.S. Geological Survey (USGS) (2021). *Significant Earthquake Archive*, <https://earthquake.usgs.gov/earthquakes/browse/significant.php>, last accessed 5/28/2021.
- U.S. Nuclear Regulatory Commission (NRC) (2012). *Confirmatory Analysis of Seismic Hazard at Diablo Canyon Power Plant from the Shoreline Fault Zone*, Research Information Letter (RIL) 12-01, September, ML 121230035.
- Valentini A., Fukushima Y., Contri P., Ono M., Sakai T., Thompson S. C., Viallet E., Annaka T., Chen R., Moss R. E. S., et al. (2021). Probabilistic fault displacement hazard assessment (PFDHA) for nuclear installations according to IAEA safety standards. *Bulletin of the Seismological Society of America*,

- 111**(5), 2661–2672, doi: 10.1785/0120210083.
- Wang, Y., Goulet C. (2022). Constraining fault displacements for strike-slip events using physics-based simulations. *Proceedings of the 12th National Conference on Earthquake Engineering*, Salt Lake City, Utah, download at <https://www.eeri.org/what-we-offer/digital-library/?lid=12791>.
- Wells D. L., Coppersmith K. J. (1993). Likelihood of surface rupture as a function of magnitude. *Seismological Research Letters*, **64**(1), 54, doi: 10.1785/gssrl.64.1.1.
- Wells D. L., Coppersmith K. J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the Seismological Society of America*, **84**(4), 974–1002, doi: 10.1785/BSSA0840040974.
- Wells D., Youngs B. (2015). Improved regression relations for earthquake source parameters. *Seismological Research Letters*, **86**(2B), 623, doi: 10.1785/0220150017.
- Wesnousky S. G. (2008). Displacement and geometrical characteristics of earthquake surface ruptures. *Bulletin of the Seismological Society of America*, **98**(4), 1609–1632, doi: 10.1785/0120070111.
- Wood S. N. (2017). *Generalized Additive Models. An Introduction with R*, Second Edition, Chapman & Hall/CRC: New York, 496 pp., doi: 10.1201/9781315370279.
- Youngs R. R., Abrahamson N., Makdisi F. I., Sadigh K. (1995). Magnitude-dependent variance of peak ground acceleration. *Bulletin of the Seismological Society of America*, **85**(4), 1161–1176, doi: 10.1785/BSSA0850041161.
- Youngs R. R., Arabasz W. J., Anderson R. E., Ramelli A. R., Ake J. P., Slemmons D. B., McCalpin J. P., Doser D. I., Fridrich C. J., Swan F. H. III, Rogers A. M., Yount J. C., Anderson L. W., Smith K. D., Bruhn R. L., Knuepfer L. K., Smith R. B., dePolo C. M., O’Leary K. W., Coppersmith K. J., Pezzopane S. K., Schwartz D. P., Whitney J.W., Olig S. S., Toro G. R. (2003). A methodology for probabilistic fault displacement hazard analysis (PFDHA). *Earthquake Spectra*, **19**(1), 191–219, doi: 10.1193/1.1542891.
- Youngs R. R., Arcos M. E., Coppersmith K. J., Coppersmith R., Kanson K. L. (2021). Fault displacement hazard: new data and modeling advances. *Seismological Research Letters*, **92**(2B), 1363, doi: [10.1785/0220210025](https://doi.org/10.1785/0220210025).

APPENDIX A Petersen et al. (2011) PFDHA Framework

Figure 2.1 in the main report illustrates the geometric elements/parameters used in the probabilistic fault displacement hazard analysis (PFDHA) framework of Petersen et al. (2011, hereafter P11), including the fault trace mapped prior to a surface rupture event (mapped fault trace), the main trace of a future surface rupture, and the coordinate system utilized to specify positions relative to the main trace. Descriptions of these parameters are given below before we introduce the basic hazard integral for annual rate of exceeding a prescribed value of displacement.

Because a future earthquake may not rupture the entire mapped fault trace, random variable s is utilized to track the along-fault-trace distance from one fixed end of the mapped fault trace to the closer end of future ruptures' main trace.

The site of interest is specified by its along-main-trace position l and its off-main-trace position Δ . Note that these two variables are defined relative to the main trace of future rupture, not the pre-event mapped fault trace. As described in Section 2.3 of the main report, the positions l and Δ are taken as the strike-parallel (u_LCP) and the strike-perpendicular (t_LCP) coordinates of the 2nd generalized coordinate system (GC2), using the main trace as the reference of GC2. In this study, the main trace of principal rupture is taken as the constructed least-cost path (LCP) of the surface rupture traces.

The footprint area of an infrastructure is considered in P11's PFDHA framework. It has a dimension z (area z^2) and it is centered at the site at a distance Δ_{site} from the main trace.

The size of an earthquake is characterized by its moment magnitude m , which is denoted as \mathbf{M} in Chapter 3 of the main report, as well as the length of main trace L taken as the length of LCP.

Given site's position Δ_{site} relative to the main trace and site's footprint area z^2 , the annual rate of the net principal displacement D exceeding a prescribed level D_0 is:

$$\lambda(D \geq D_0 | \Delta_{site}, z) = \alpha(m_0) \int_{m,s} f_{M,S}(m, s) P[sr \neq 0 | m] P_2[D \neq 0 | z, \Delta_{site}, sr \neq 0] P_1[D \neq 0 | z, sr \neq 0] P[D \geq D_0 | l/L, m, D \neq 0] \, dmds \quad (A.1)$$

where:

- $\alpha(m_0)$ the annual rate of earthquakes above a minimum magnitude (m_0) of engineering significance or of a characteristic earthquake magnitude on a fault source;
- $f_{M,S}(m, s)$ a function characterizing the joint density function of earthquake magnitude (m) and location (s) of future ruptures associated with the pre-event mapped fault trace;
- $P[sr \neq 0 | m]$ the probability of having surface rupture given that a magnitude m earthquake occurs. Additional discussions of this probability term are given below;
- $P_2[D \neq 0 | z, \Delta_{site}, sr \neq 0]$ the probability that the main trace of a future surface rupture intersects the footprint area z^2 . This probability term was defined per Chen and Petersen (2019) to account for the uncertainty in the location of future surface rupture relative to the pre-event mapped fault trace. If the site falls outside the ends of main trace, $P_2 = 0$; otherwise, P_2 is computed by integrating $f_R(r)$ from $r = \Delta_{site} - \frac{z}{2}$ to $r = \Delta_{site} + \frac{z}{2}$ (see Figure 4 of Chen and Petersen, 2019). The function $f_R(r)$ is the probability density function of r , which is the shortest distance from an arbitrary point on surface rupture traces to the pre-event mapped fault trace (see Figure A.1 and Figure 3 of P11). Note that the (non-random) distance Δ_{site} was improperly denoted as r in Figure 4 of Chen and Petersen (2019). An empirically derived $f_R(r)$ was given in Tables 2 and 3 of P11. Additional discussions of $f_R(r)$ are given below;
- $P_1[D \neq 0 | z, sr \neq 0]$ the probability that the footprint area z^2 does not completely fall within a surface rupture gap (by definition, principal displacement within a rupture gap is zero), given that main trace intersects the footprint area;
- $P[D \geq D_0 | l/L, m, D \neq 0]$ the probability that principal displacement D exceeds the prescribed level D_0 , given the conditions that the surface rupture event is of magnitude m , its main trace intersects site's footprint area, site's footprint area does not completely fall within a rupture gap on the main trace, and site's normalized along-main-trace position is l/L . This probability is computed as the complementary cumulative distribution function (CDF) of the probability distribution of principal displacement, which is prescribed by a fault displacement model (FDM).

Additional discussions of the above terms are given below. Typically, in hazard analysis, it is assumed that a fault may rupture: (1) repetitively with similar magnitude earthquakes (over the same location) (i.e., the characteristic earthquake recurrence model), (2) with a sequence of earthquakes described by an exponential Gutenberg-Richter magnitude-frequency distribution, and (3) with earthquakes that follow a combination of the characteristic and Gutenberg-Richter magnitude-frequency distributions. These magnitude-frequency distributions are determined from examination of historical seismicity, consideration of the physical constraints on the length or area of the fault, complexity of the fault along strike, and crustal rheology properties along the fault. Earthquake magnitude is estimated from rupture length, rupture area, or magnitude of historical earthquakes with uncertainty. Observations indicate that faults do not always rupture along the entire mapped length (e.g., the 1868 **M** 6.8 Hayward and 1933 **M** 6.4 Long Beach, California, earthquakes), and rupture may also jump to adjacent faults (e.g., 1992 **M** 7.2 Landers, California, earthquake). The joint probability density function (PDF), $f_{M,S}(m,s)$, enables consideration of the potential for partial rupture occurring over various lengths of the fault. The range of s is from zero to the total fault length minus the rupture length. In most PSHA literature, magnitude-frequency distribution is denoted as $f_M(m)$ (not to be confused with the magnitude-scaling term $f_M(\mathbf{M})$ used in FDM), and partial rupture is factored in a separate source to site distance PFD. P11's PFDHA framework uses the joint PDF, $f_{M,S}(m,s)$, to simplify the hazard equation.

$P_2[D \neq 0 | z, \Delta_{site}, sr \neq 0]$ is computed by integrating the PDF $f_R(r)$ over the r range from $\Delta_{site} - z/2$ to $\Delta_{site} + z/2$; see Figure 4 of Chen and Petersen (2019). The PDF, $f_R(r)$, is defined to allow inclusion of uncertainty (and variability) in rupture location relative to the mapped fault trace. In P11, location uncertainty is quantified by the PDF of distances r measured from the locations of coseismic surface rupture mapped following an earthquake to the nearest fault trace mapped prior to the event. The distance data indicate an approximately normal distribution that centers on the mapped fault (P11; Chen et al., 2013). In PFDHA, location uncertainty translates into variation of calculated principal-fault displacement in the direction perpendicular to fault strike. Consequently, the calculated principal-fault displacement along a line across the fault trace also resembles the shape of a normal distribution (often with truncation to save computational time), exhibiting a bell-shaped profile centered on the mapped fault (P11, 2011; Chen and Petersen, 2019). The standard deviation of $f_R(r)$ depends on the mapping accuracy of the pre-event mapped fault trace (categorized as accurately located, approximately located, inferred, or concealed) and fault complexity (simple versus complex). In P11's formulation, variable r is used for both location uncertainty and the closest distance from the site to the mapped fault trace. In the current report, to avoid confusion, we use r to denote location uncertainty only and use Δ_{site} to denote the closest distance from the site to the mapped fault.

A fault typically is a complex zone that is made of coalescing faults and shears. The surface ruptures may not occur on the same trace during subsequent earthquakes, and faults may evolve through time, creating new fault traces. This aleatory variability contributes to the random location of surface rupture of future earthquakes and is not presently well-quantified. In addition, mapped

fault traces have inaccuracies that translate into an epistemic uncertainty in the location of future ruptures.

In P11, the PDF for r includes both aleatory and epistemic components of location uncertainty. Ideally, these two types of uncertainties should be treated separately in the PFDHA numerical algorithm. However, currently, there are no data that distinguish them. To separate these uncertainties, each one needs to be defined quantitatively. Paleoseismic data (e.g., Liu-Zeng et al., 2004, 2006; Dawson et al., 2003) can be used to quantify the natural variability in rupture location. However, such data need to be systematically collected and statistically analyzed before they can be used in PFDHA (Chen et al., 2013).

$P[sr \neq 0|m]$ accounts for the probability that an earthquake rupture on a mapped fault reaches the surface. For example, the 1989 **M** 6.9 Loma Prieta, California, and the 2002 **M** 6.7 Nenana Mountain, Alaska, earthquake ruptures did not extend to the surface and, therefore, did not present a surface fault displacement hazard. This probability can be computed using either simulation or empirical models (Youngs et al., 2003). P11 used the empirical formulation developed by Wells and Coppersmith (1993). Their equation for calculating the probability of surface rupture is given by a logistic regression model (commonly applied when the dependent variable is dichotomous) that provides the probability of surface rupture conditioned on magnitude m :

$$P[sr \neq 0|m] = \frac{e^{(a+bm)}}{1 + e^{(a+bm)}} \quad (\text{A.2})$$

where sr is a binary variable, with $sr \neq 0$ representing the occurrence of surface rupture, and constants a and b are regression coefficients that depend on fault type. For example, a and b are -12.51 and 2.053, respectively, for strike-slip faults, implying probabilities of 87% that an **M** 7 earthquake will rupture to the surface and 95% that an **M** 7.5 earthquake will rupture to the surface. Other relations may also be derived from local/regional data of surface rupturing events (e.g., Moss and Ross, 2011, for reverse faults).

The term $P_I[D \neq 0|z, sr \neq 0]$ represents the probability of non-zero displacement given that the main trace of surface rupture passes through the footprint area z^2 . It accounts for gaps of surface rupture directly along the main trace. For large through-going strike-slip faults in California, it is reasonable to assume that $P_I[D \neq 0|z, sr \neq 0]$ is 1.0. In other cases, this assumption may not be appropriate, particularly for less mature faults or faults with complex geometry. Many mapped principal ruptures in the FDHI database are fragmented, and gaps with no measurable surface displacement are observed at all scales. One example of events with highly fragmented surface rupture is the Kumamoto earthquake.

$P[D \geq D_0|l/L, m, D \neq 0]$ is the conditional probability that principal displacement D exceeds a prescribed D_0 . This exceedance probability is computed as the complementary CDF of the adopted probability distribution of principal displacement. For example, in P11, the exceedance probability is obtained by using the complementary CDF of log-normal distribution. For the preferred model of this study, it is computed using the complementary CDF of negative exponentially modified Gaussian (nEMG) distribution given in Appendix C.

Each PDF or probability term discussed above is a component model in the P11 PFDHA framework. This framework and data needed to develop empirical models for each component are summarized in Figure A.1 for principal displacement. The FDMs discussed in Chapter 3 of the main report are needed in the calculation of the conditional exceedance probability for principal displacement (i.e., the $P[D \geq D_0|l/L, m, D \neq 0]$ term in Figure A.1).

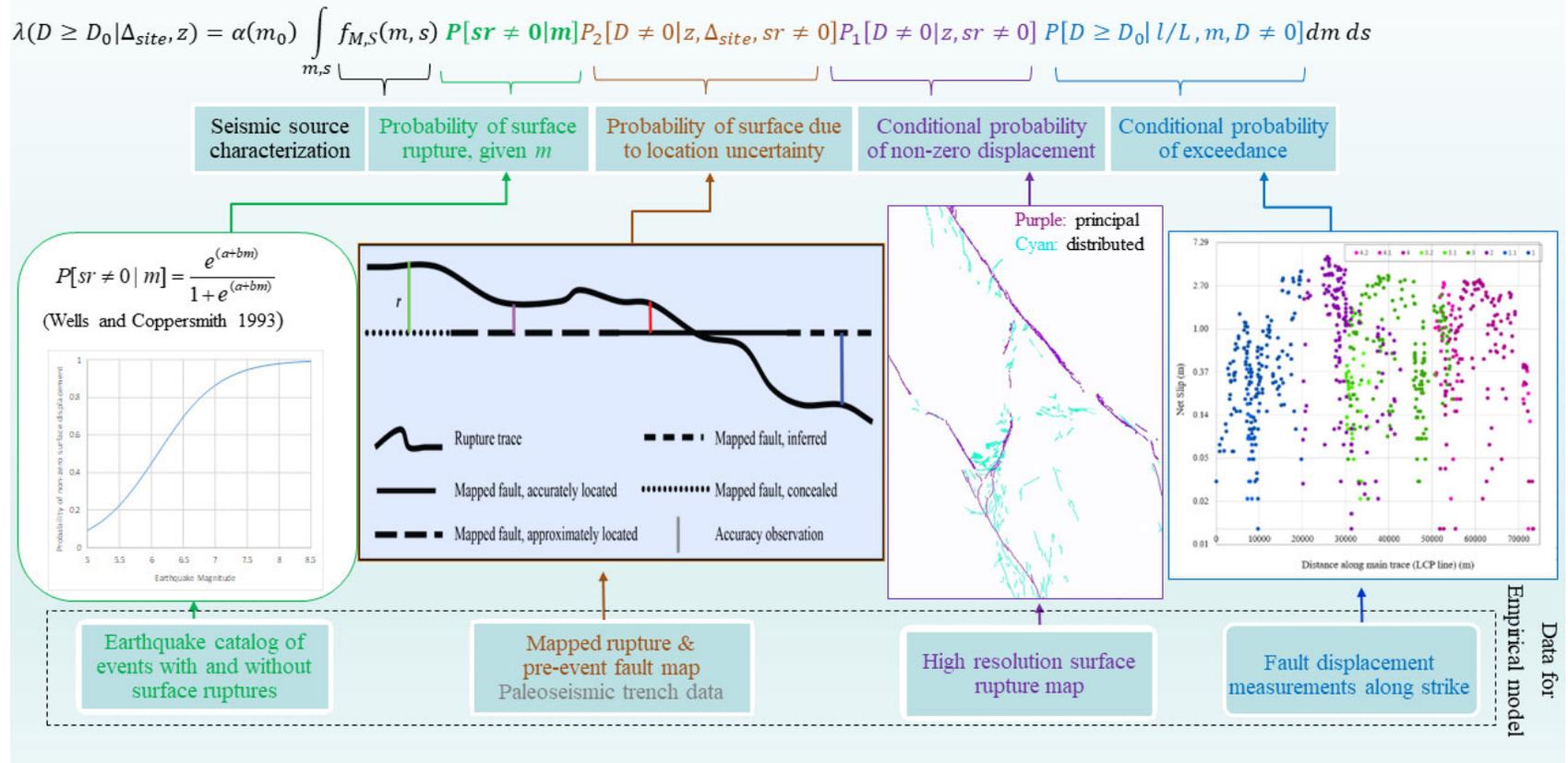
Except for the $P[D \geq D_0|l/L, m, D \neq 0]$, other components in the P11 framework are not updated in this study. The Fault Displacement Hazard Initiative (FDHI) database (Sarmiento et al., 2021) can be used to update some of these models in the future. As an example, the FDHI database can be used to update the $P_l[D \neq 0|z, sr \neq 0]$ term. However, the FDHI database does not contain data needed to define the conditional probability of surface rupture given magnitude or the uncertainty in the location of future surface rupture.

REFERENCES

- Chen R., Petersen M. D. (2019). Improved implementation of rupture location uncertainty in fault displacement hazard assessment. *Bulletin of the Seismological Society of America*, **109**(5), 2132–2137, doi: 10.1785/0120180305.
- Chen R., Dawson T. E., Wills C. J. (2013). *Quantifying Surface Fault Rupture Location Uncertainties for Lifeline Crossings*, Final Technical Report, U.S. Geological Survey National Earthquake Hazards Reduction Program, Award Number G11AP20040, 51 pp.
- Dawson T. E., McGill S. F., Rockwell T. K. (2003). Irregular recurrence of paleoearthquakes along the central Garlock fault near El Paso Peaks, California. *Journal of Geophysical Research: Solid Earth*, **108**(B7), 2356, doi: 10.1029/2001JB001744 .
- Liu-Zeng J., Klinger Y., Sieh K., Rubin C. (2004). Six similar sequential ruptures of the San Andreas Fault, Carrizo Plain, California. *Geology*, **32**(8), 649–652, doi: 10.1130/G20478.1.
- Liu-Zeng J., Klinger Y., Sieh K., Rubin C., Seitz G. (2006). Serial ruptures of the San Andreas Fault, Carrizo Plain, California, revealed by three-dimensional trench excavations. *Journal of Geophysical Research: Solid Earth*, **111**(B2), B02306, doi: 10.1029/2004JB003601.
- Moss R. E. S., Ross Z. E. (2011). Probabilistic fault displacement hazard analysis for reverse faults. *Bulletin of the Seismological Society of America*, **101**(4), 1542–1553, doi: 10.1785/0120100248.
- Petersen M. D., Dawson T. E., Chen R., Cao T., Wills C. J., Schwartz D. P., Frankel A. D. (2011). Fault displacement hazard for strike-slip faults. *Bulletin of the Seismological Society of America*, **101**(2), 805–825, doi: 10.1785/0120100035.
- Sarmiento A., Madugo D., Bozorgnia Y., Shen A., Mazzoni S., Lavrentiadis G., Dawson T., Madugo C., Kottke A., Thompson S., Baize S., Milliner C., Nurminen F., Boncio P., Visini F. (2021). *Fault*

Displacement Hazard Initiative Database, UCLA B. John Garrick Institute for the Risk Sciences, Report GIRS-2021-08, Revision 3 dated 19 July 2022, 97 pp., doi: 10.34948/N36P48.

- Wells, D. L., Coppersmith K. J. (1993). Likelihood of surface rupture as a function of magnitude. *Seismological Research Letters*, **64**(1), 54, doi: 10.1785/gssrl.64.1.13.
- Wells D. L., Coppersmith K. J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the Seismological Society of America*, **84**(4), 974–1002, doi: 10.1785/BSSA0840040974.
- Wells D., Youngs B. (2015). Improved regression relations for earthquake source parameters. *Seismological Research Letters*, **86**(2B), 623, doi: 10.1785/0220150017.
- Youngs R. R., Arabasz W. J., Anderson R. E., Ramelli A. R., Ake J. P., Slemmons D. B., McCalpin J. P., Doser D. I., Fridrich C. J., Swan F. H. III, Rogers A. M., Yount J. C., Anderson L. W., Smith K. D., Bruhn R. L., Knuepfer L. K., Smith R. B., dePolo C. M., O’Leary K. W., Coppersmith K. J., Pezzopane S. K., Schwartz D. P., Whitney J. W., Olig S. S., Toro G. R. (2003). A methodology for probabilistic fault displacement hazard analysis (PFDHA). *Earthquake Spectra*, **19**(1), 191–219, doi: 10.1193/1.1542891.



APPENDIX B Data Selection Tables

This appendix presents four data selection tables referred to in Chapter 2. These include:

Table B.1. Strike-slip events in the Fault Displacement Hazard Initiative (FDHI) database

Table B.2. Rupture dataset selection and rationale

Table B.3. Explanation of data quality indices in the Fault Displacement Hazard Initiative (FDHI) database and usage recommendation (slightly simplified from Sarmiento et al., 2021)

Table B.4. Measurement dataset selection and rationale

Table B.1. Strike-slip events in the Fault Displacement Hazard Initiative (FDHI) database

EQ_ID ³	Eq_name ⁴	Region	Year	Mw ⁵	Number of Datasets	
					Rupture	Measurement
4	Balochistan	Pakistan	2013	7.70		2
6	Borrogo ¹	California	1968	6.63		
36	ChalfantValley	California	1986	6.19		2
21	Darfield	New Zealand	2010	7.00	2	3
10	Denali ²	Alaska	2002	7.90		3
11	Duzce ¹	Turkey	1999	7.14	2	4
57	ElmoreRanch	California	1987	6.22		
2	HectorMine ¹	California	1999	7.13		4
28	Imperial1979 ²	California	1979	6.53		
7	Imperial1940 ¹	California	1979	6.95		2
5	Izmit_Kocaeli ¹	Turkey	1999	7.51		2
9	Kobe ¹	Japan	1995	6.90		2
17	Kumamoto	Japan	2016	7.00		
1	Landers ¹	California	1992	7.28		2
55	Luzon ²	Philippines	1990	7.70		
13	Napa	California	2014	6.00		
65	Neftegorsk	Russia	1995	7.00		
29	Parkfield1966	California	1966	6.19		
22	Parkfield2004	California	2004	6.00		
42	Ridgecrest1	California	2019	6.40	2	
43	Ridgecrest2	California	2019	7.10	2	
53	SanMiguel	Mexico	1956	6.80		
8	SuperstitionHills ¹	California	1987	6.54		
14	Yushu	China	2010	6.90		2
54	Yutian	China	2014	6.90		2
37	Zirkuh	Iran	1997	7.20		
67	Kunlun_Kokoxili	Northern Tibet	2001	7.80		2
71	Palu	Indonesia	2018	7.50		3
75	YeniceGonen	Turkey	1953	7.30		
32	GalwayLake	California	1975	5.20		
70	HomesteadValley	California	1979	5.20		
15	Hualien	Taiwan	2018	6.40		
62	IzuOshima	Japan	1978	6.60		
61	IzuPeninsula	Japan	1974	6.50		
58	Pisayambo	Ecuador	2010	5.00		

¹Used in the Petersen et al. (2011) study; ²used in Petersen et al. (2011) principal displacement model, with data from Wesnousky (2008); ³Earthquake identification number in FDHI database; ⁴Earthquake name in FDHI database; ⁵See FDHI database (Sarmiento et al., 2021) for magnitude type (mostly Mw with a few exceptions).

Table B.2. Rupture dataset selection and rationale

Event	<i>RUP_DS_ID</i> ¹		Reasons
	Selected	Excluded	
Darfield	103	80	FDHI recommendation
Duzce	43	36	More complete spatial coverage, companion measurement dataset
Ridgecrest1	132	145	FDHI recommendation
Ridgecrest2	132	145	FDHI recommendation

¹Rupture dataset identification number in the Fault Displacement Hazard Initiative (FDHI) database (Sarmiento et al., 2021).

Table B.3. Explanation of data quality indices in the Fault Displacement Hazard Initiative (FDHI) database (October 2020 release) and usage recommendation (slightly simplified)¹

qual_code	net_flag	Explanation	Recommendation
1	Keep	Good data	Good data
2000	Check	Same site (or obviously same site) measurement from another dataset	Check available alternatives
2001	Check	Nearby measurement from another dataset	Check available alternatives
9000	Toss	Other measurement of same feature is more complete	Bad data
9001	Toss	No measurement data	Bad data
9002	Toss	Incomplete measurement, significant lateral slip unaccounted for	Bad data
9003	Toss	Incomplete measurement, significant vertical slip unaccounted for	Bad data
9004	Toss	Measurement likely erroneous	Bad data
9005	Toss	Location likely erroneous	Bad data
9006	Toss	Deformation likely not tectonic	Bad data
3000	Check	Incomplete measurement, lateral slip component might be missing	Check measurement completeness/quality
3001	Check	Incomplete measurement, vertical slip component might be missing	Same as above
3002	Check	Measurement might be minimum	Same as above
3003	Check	Measurement might be maximum	Same as above
3004	Check	Author quality is low	Same as above
3005	Check	Deformation might not be tectonic	Same as above
4000	Check	Location might be erroneous	Check for possible error
4001	Check	Measurement might be erroneous	Same as above
5000	Check	Measurement technique might mis-estimate vertical slip component	Check measurement technique

¹Reproduced from FDHI database documentation, a worksheet titled “quality_code_explanation” in a Microsoft Excel file named “FDHI_Database_ph16_rev2_FIELDDEFINITIONS_20201005.”

Table B.4. Measurement dataset selection and rationale

Event	<i>PT_DS_ID</i> ¹		Rationale
	Selected	Excluded	
Balochistan	23	75	Companion of the rupture dataset
ChalfantValley	126	125	Larger number of high-quality principal displacement measurements
Darfield	78	77, 79	The most complete spatial coverage, largest number of high-quality principal displacement measurements
Denali	39	40, 90	Largest number of high-quality principal displacement measurements
Duzce	43	37, 38, 144	A much larger dataset, largest number of high-quality principal displacement measurements, companion of the selected rupture dataset
Hector	2	3, 6, 99	Largest number of high-quality principal displacement measurements, an updated version of 6. The other two datasets are mostly ranked as “total”
Hualien	62	61	All measurements in 61 are ranked as “total”
Imperial1940	96	162	162 has very limited spatial coverage and other reasons discussed in the text
Izmit_Kocaeli	6	144	Larger number of good quality principal displacement measurements, companion of rupture dataset
Kobe	86	6	Larger number of good quality principal displacement measurements
Landers	6	3	Larger number of good quality principal displacement measurements, 3 is mostly ranked as “total”
Yushu	58	57	Larger number of good quality principal displacement measurements
Yutian	146	147	Larger number of good quality principal displacement measurements
Kunlun_Kokixili	52	93	Field-based measurements, better spatial coverage; 93 covers only central ≈ 100 km of rupture, is satellite-based, general high slip than field-based measurements
Palu	171	169, 170	Companion of rupture dataset, largest number of high-quality measurements, better spatial coverage

¹These are measurement point identification numbers in the Fault Displacement Hazard Initiative (FDHI) database (Sarmiento et al., 2021).

APPENDIX C Implementation of Probability Distributions

C.1 SKEW-NORMAL DISTRIBUTION

C.1.1 PDF and CDF

The probability density function (PDF) f_Y of skew-normal variate $Y \sim SN(\mu, \sigma, \nu)$ is reproduced from Table 18.9 of Rigby et al. (2020),

$$f_Y(y) = \frac{2}{\sigma} \phi(z) \cdot \Phi(\nu z)$$
$$z = \frac{y - \mu}{\sigma}$$

where ϕ and Φ are the PDF and the cumulative distribution function (CDF) of standard normal distribution. Rigby et al. (2020) does not provide an analytic expression of the CDF (F_Y) of skew-normal distribution; instead, F_Y in R's package *gamlss* (Stasinopoulos et al., 2017) is computed by numerical integration of f_Y . Alternatively, an analytic expression of F_Y in terms of Owen's T function $T(z, \nu)$ (Owen, 1956) is given in Azzalini and Capitanio (2014; Equation 2.37),

$$F_Y(y) = \Phi(z) - 2 \cdot T(z, \nu)$$

C.1.2 R implementation

Skew-normal distribution is implemented in R's packages *sn* (Azzalini and Capitanio, 2014) and *gamlss*. The CDF in *gamlss* package is computed using numerical integration of the PDF, while CDF in *sn* package uses the alternative expression that involves Owen's T function.

C.1.3 FORTRAN implementation

A FORTRAN subroutine for the CDF of skew-normal distribution was coded in April 2022 and incorporated into California Geological Survey's (CGS's) PFDHA code, which was then used in the example hazard calculations described in Chapter 4. The subroutine *psn.f* (listed below) is based on the above expression of F_Y in which the Owen's T function is computed using the fast and accurate algorithm of Patefield and Tandy (2000). We did not code the PDF of skew-normal distribution in FORTRAN because it is not needed in PFDHA.

C.1.4 SOURCE CODES

C.1.4.1 *psn.f*

FORTRAN code *psn.f* can be downloaded from <https://www.conservation.ca.gov/cgs/pfdha>. No source code listing is provided in this report.

C.2 SKEW-T DISTRIBUTION

From Table 18.25 of Rigby et al. (2020), the PDF of skew- t distribution is

$$f_Y = \frac{2}{\sigma} f_{Z_1}(z) F_{Z_2}(\omega)$$
$$z = (y - \mu)/\sigma$$
$$\omega = \nu z \sqrt{(\tau + 1)/(\tau + z^2)}$$

where f_{Z_1} is the PDF of Student's- t with τ degree of freedom ($Z_1 \sim t_\tau$) and $F_{Z_2}(\omega)$ is the CDF of Student's- t with $\tau + 1$ degree of freedom ($Z_2 \sim t_{\tau+1}$). Analytic expression of skew- t 's CDF is not provided in Rigby et al. (2020) and Azzalini and Capitanio (2014).

C.2.1 R implementation

Skew- t distribution is implemented in R's packages *sn* and *gamlss*. The CDF in both *sn* and *gamlss* package is computed using numerical integration of the PDF.

C.2.2 FORTRAN implementation

We coded the PDF of skew- t distribution in FORTRAN (*pst.f*). The same FORTRAN code also computes the CDF of skew- t distribution by numerical integration of its PDF. This subroutine was written in November 2002 and has yet to be incorporated into CGS's PFDHA code.

C.2.3 Source codes

FORTRAN code *pst.f* can be downloaded from <https://www.conservation.ca.gov/cgs/pfdha>. No source code listing is provided in this report.

C.3 nEMG DISTRIBUTION

C.3.1 PDF and CDF

In this section, we present the analytical expressions of the PDF (f_Y) and CDF (F_Y) of negative exponentially modified Gaussian (nEMG) variate $Y = (G - E) \sim nEMG(\mu, \sigma, \nu)$. We note that, because $X = -Y = (-G) + E$, X is an exGaussian variate whose Gaussian component has a mean of $(-\mu)$, $X \sim exGAUS(-\mu, \sigma, \nu)$. PDF f_Y and CDF F_Y presented herein are derived by taking advantages of the relation $-Y = X$ and the analytic expressions of X 's PDF (f_X) and CDF (F_X).

C.3.1.1 PDF and CDF of exGaussian Variate

The distribution functions of exGaussian variate X , as implemented in R's *gamlss* package, has the following analytic expressions.

$$f_X(x | \mu, \sigma, \nu) = \begin{cases} \frac{1}{\nu} e^{-\frac{(z + \frac{\sigma^2}{2\nu})}{\nu}} \Phi\left(\frac{z}{\sigma}\right), & \nu > 0.05\sigma \\ \phi\left(\frac{x - \mu}{\sigma}\right), & \nu \leq 0.05\sigma \end{cases}$$

$$z = x - \mu - \frac{\sigma^2}{\nu}$$

$$F_X(x | \mu, \sigma, \nu) = \begin{cases} \Phi\left(\frac{x - \mu}{\sigma}\right) - \Phi\left(\frac{z}{\sigma}\right) e^{\frac{(\mu + \frac{\sigma^2}{\nu})^2 - \mu^2 - 2x\frac{\sigma^2}{\nu}}{2\sigma^2}}, & \nu > 0.05\sigma \\ \Phi\left(\frac{x - \mu}{\sigma}\right), & \nu \leq 0.05\sigma \end{cases}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the PDF and CDF of the standard normal distribution.

C.3.1.2 PDF and CDF of nEMG Variate

Because $-Y = X$, the PDF of Y is related to the PDF of X as

$$f_Y = f_X(-x).$$

By substituting $-y$ for x and $-\mu$ for μ in the expression for f_X , we have

$$f_Y(y | \mu, \sigma, \nu) = \begin{cases} \frac{1}{\nu} e^{-\frac{(z + \frac{\sigma^2}{2\nu})}{\nu}} \Phi\left(\frac{z}{\sigma}\right), & \nu > 0.05\sigma \\ \phi\left(\frac{-y + \mu}{\sigma}\right), & \nu \leq 0.05\sigma \end{cases}$$

$$z = -y + \mu - \frac{\sigma^2}{\nu}$$

To obtain the PDF of Y , we use again $-Y = X$ and the property

$$F_Y(y) = P(Y \leq y) = P(-X \leq y) = P(X \geq -y) = 1 - P(X \leq -y) = 1 - F_X(-y)$$

and we get

$$F_Y(y | \mu, \sigma, \nu) = 1 - F_X(-y | -\mu, \sigma, \nu).$$

Substituting $-y$ for x and $-\mu$ for μ in the expression for F_X , we have

$$F_Y(y | \mu, \sigma, \nu) = 1 - \begin{cases} \left\{ \Phi\left(\frac{-y + \mu}{\sigma}\right) - \Phi\left(\frac{z}{\sigma}\right) e^{\frac{(-\mu + \frac{\sigma^2}{\nu})^2 - \mu^2 + 2y\frac{\sigma^2}{\nu}}{2\sigma^2}} \right\}, & \nu > 0.05\sigma \\ \Phi\left(\frac{-y + \mu}{\sigma}\right), & \nu \leq 0.05\sigma \end{cases}$$

C.3.2 R implementation

There are two versions of R implementations for PDF and CDF. The first version is based on the relation between $f_Y(y)$ and $f_X(x)$ for PDF (*dnEMG.r*), and between $F_Y(y)$ and $F_X(x)$ for CDF (*pnEMG.r*). The second version (*nEMG.r*) is based on the analytic expressions of $f_Y(y)$ and $F_Y(y)$ presented above.

C.3.3 FORTRAN implementation

A FORTRAN subroutine for the CDF of Y ($F_Y(y|\mu, \sigma, \nu)$) was coded in May 2022 and incorporated into CGS's PFDHA code, which was then used in the example hazard calculations described in Chapter 4. The subroutine *pnEMG.f* (listed below) is based on the relation between $F_Y(y)$ and $F_X(x)$ described above. We did not code the PDF of nEMG distribution because it is not needed in PFDHA.

C.3.4 Source codes

C.3.4.1 dnEMG.r

```
dnENG <- function (x, mu = 5, sigma = 1, nu = 1, log = FALSE) {
  if (any(sigma <= 0))
    stop(paste("sigma must be greater than 0 ", "\n", ""))
  if (any(nu <= 0))
    stop(paste("nu must be greater than 0 ", "\n", ""))
  logfy <- gamlss::dexGAUS(-x, mu = -mu, sigma = sigma, nu = nu, log =
T)
  if (log == FALSE)
    fy <- exp(logfy)
  else fy <- logfy
  fy
}
```

NOTE: R function `gamlss::dexGAUS` is the PDF of exGaussian distribution

C.3.4.2 pnEMG.r

```
pnENG <- function (q, mu = 5, sigma = 1, nu = 1, lower.tail = TRUE,
log.p = FALSE) {
  if (any(sigma <= 0))
    stop(paste("sigma must be greater than 0 ", "\n", ""))
  if (any(nu <= 0))
    stop(paste("nu must be greater than 0 ", "\n", ""))

  cdf <- 1 - gamlss::pexGAUS(-q, -mu, sigma, nu, lower.tail =
lower.tail, log.p = F)
  if (lower.tail == TRUE)
    cdf <- cdf
}
```

```

else cdf <- 1 - cdf

if (log.p == FALSE)
  cdf <- cdf
else cdf <- log(cdf)
cdf
}

```

NOTE: R function `gamlss::pexGAUS` is the CDF of exGaussian distribution.

C.3.4.3 nEMG.r

```

nEMG <- function(y, mu, sig, nu) {
  z <- -y + mu - sig^2/nu
  if(mu > 0.05 * sig) {
    f_y <- 1/nu * exp(-(z+sig^2/(2*nu))/nu) * pnorm(z/sig) # eq 1.4
    F_y <- 1 - (pnorm((-y + mu) / sig) - pnorm(z/sig) * exp((( -mu + sig^2/nu)^2 - mu^2 + 2 * y * sig^2 /
nu) / (2 * sig^2)))
  } else {
    f_y <- dnorm((-y + mu) / sig)
    F_y <- pnorm((-y + mu) / sig)
  }
  return(list(f_y = f_y, F_y = F_y))
}

```

NOTE: R functions `dnorm` and `pnorm` are the PDF and CDF of normal distribution, respectively.

C.3.4.4 pnEMG.f

FORTTRAN code `pnEMG.f` can be downloaded from <https://www.conservation.ca.gov/cgs/pfdha>. No source code listing is provided in this report.

REFERENCES

Azzalini A., Capitanio A. (2014). *The Skew-Normal and Related Families*, Institute of Mathematical Statistics Monographs, Cambridge University Press: New York, 262 pp., doi: 10.1017/CBO9781139248891.

Owen D. B. (1956). Tables for computing bivariate normal probabilities. *Annals of Mathematical Statistics*, 27(4), 1075–1090, doi: 10.1214/aoms/1177728074.

Patefield M., Tandy D. (2000). Fast and accurate calculation of Owen's T-function. *Journal of Statistical Software*, 5(5), 1–25, doi: 10.18637/jss.v005.i05.

Rigby R. A., Stasinopoulos M. D., Heller G. Z., Bastiani F. De (2020). *Distribution for Modelling Location, Scale, and Shape: Using GAMLSS in R*, Chapman & Hall/CRC: New York, 588 pp., doi: 10.1201/9780429298547.

Stasinopoulos M. D., Rigby R. A., Heller G. Z., Voudouris V., Bastiani F. D. (2017). *Flexible Regression and Smoothing: Using GAMLSS in R*, Chapman & Hall/CRC: New York, 571 pp., doi: 10.1201/b21973.

APPENDIX D Tier Classification of FDHI Principal Displacement Data

In Section 3.2.2 the level of data non-normality is reduced by excluding a small number of small-displacement observations whose residuals are deemed incompatible with the assumed normal distribution. Such trimming is achieved by classifying displacement data in a particular earthquake into two tiers. The tier-1 class, which consists of the majority of observations, is intended to be the subset of displacements that follows roughly a hypothetical normal distribution. The tier-2 class, which consists of a small number of small-displacement observations, is the complement of the tier-1 subset. The method utilized to define the boundary between tier-1 and tier-2 is described in this Appendix.

D.1 TIER BOUNDARY

Tier classification of displacement data is actually based on the natural logarithm of displacement but, for the sake of brevity, we will still refer to it as displacement. In this study, tier boundary is taken as the estimated lower bound of the hypothetical normal population. The estimation of this lower bound, per earthquake, is outlined as follows:

1. The 0.997 ($y_{0.997}$) and the 0.5 ($y_{0.5}$) quantiles of fault displacements are modeled separately by the same function $f_{l2L}(l2L)$ but with different coefficient values. Variable $l2L$ is the normalized along-trace position of a displacement measurement point; see Section 3.2.1.1 of the main report. More details on the function $f_{l2L}(l2L)$ are given below.
2. Quantile regression (Koenker, 2005) is utilized to infer the model coefficients for $y_{0.997}$ and $y_{0.5}$ separately using the displacement data in an individual earthquake. The estimated models provide estimates of $y_{0.997}$ and $y_{0.5}$ for points along the main trace of that earthquake.

- Quantile regression is a robust and distribution-free regression method. Hence, it does not matter whether displacement data follow a normal distribution or not.
3. The estimated 0.997 and 0.5 quantiles are assumed to be the same as the 0.997 and 0.5 quantiles of the hypothetical normal distribution of the tier-1 displacement.
 4. The lower bound of displacements taken from the hypothetical normal distribution can be ideally defined by the 0.003 quantile, or $y_{0.5} - (y_{0.997} - y_{0.5})$. However, we place a floor on $(y_{0.997} - y_{0.5})$ to help stabilize the along-length variation of the lower bound when the estimated $(y_{0.997} - y_{0.5})$ interval varies widely over the rupture length. We choose the floor fl as the 75th percentile of the $(y_{0.997} - y_{0.5})$ values predicted for 101 locations equally spaced along the rupture length. Such floor also helps prevent excessive assignment of data to the tier-2 class. With this floor, the lower bound is defined by the curve of $y_{0.5} - \max(y_{0.997} - y_{0.5}, fl)$.

An example of tier boundary calculation is given on Figure D.1.

D.2 FUNCTIONAL FORMS FOR $l2L$ -DEPENDENCE

Preliminary evaluations reveal that, for earthquakes with sparse measurements, quantile estimation (and hence the estimated lower bound of tier-1 displacement) would be sensitive to the selected functional form of $f_{l2L}(l2L)$. To account for such sensitivity, we adopt eight different functional forms and synthesize the classification results into a single final classification of a data point.

In the following, the q quantile of $\ln(D)$ is denoted as y_q , where q is either 0.997 or 0.5. The set of $f_{l2L}(l2L)$ functions are summarized below.

1. Elliptical (2-parameter; symmetric with respect to $l2L = 0.5$): $\ln(y_q) = a + b \left(\sqrt{1 - \frac{(l2L-0.5)^2}{0.5^2}} - 1 \right)$
 - This f_{l2L} function, which is symmetric with respect to $l2L = 0.5$, is the same function as used in this study to model along-trace variation of the location parameter of assumed probability distribution; see Chapter 3 of the main text.
2. Elliptical + linear $l2L$ term (3-parameter): $\ln(y_q) = a + b \left(\sqrt{1 - \frac{(l2L-0.5)^2}{0.5^2}} - 1 \right) + c (l2L - 0.5)$

- This functional form is an extension of the elliptical form to incorporate asymmetry in along-trace variation. See also Section 3.6.2 of the main text.
3. Quadratic (3-parameter): $\ln(y_q) = a + b (l2L) + c (l2L)^2$
 - Note that it is not the same function as used in the quadratic model of Petersen et al. (2011; P11) (P11 uses $l2L_f$ as the covariate, which results in a symmetric function with respect to $l2L = 0.5$).
 4. Cubic (3-parameter): $\ln(y_q) = a + b c (l2L) - b (c + 1) l2L^2 + b (l2L)^3$
 - This cubic functional form is specialized to ensure that y_q at both edges of the main trace is equal to a .
 5. Modified beta PDF (4-parameter): $\ln(y_q) = e^c l2L^{e^a} (1 - l2L)^{e^b} + d$
 - This functional form was first proposed by Dr. Nico Kuehn and it was adopted and used in the early stage of our investigation to model the mean of $\ln(D)$ under normality assumption.
 - The exponentiations in e^a , e^b , and e^c is a way to impose positive-value constraint on the model coefficients a , b , and c .
 6. B-spline function of three degrees of freedom, as implemented by the $bs(df=3)$ function of package ‘*splines*’ in R (Hastie, 1992)
 7. B-spline function of five degrees of freedom, as implemented by the $bs(df=5)$ function of package ‘*splines*’ in R
 8. B-spline function of seven degrees of freedom, as implemented by the $bs(df=7)$ function of package ‘*splines*’ in R

Except for the two-parameter elliptical functional form, all of the above functional forms do not necessarily peak at $l2L = 0.5$.

D.3 SYNTHESIS OF TIER CLASSIFICATIONS

For small-displacement data, the outcomes of tier classifications using the above eight different functional forms of $f_{l2L}(l2L)$ (Figure D.2) are not always consistent. Such inconsistency poses a challenge to obtaining a single classification for a data point. As an example, when a data point is classified as tier-2 by 2 of the 8 functional forms, is it appropriate to ignore the two tier-2 classifications and treat that data point as tier-1? What is the number of times of tier-2 classifications for a data point to be given a final classification of tier 2? We could not find a

criterion that applies to all earthquakes, because the tier-classification inconsistency appears to depend on the number of measurements in a particular earthquake and the complexity of displacement's along-trace variation in that earthquake. We resort to using an earthquake-specific criterion that is determined by manually examining the classification results for each individual earthquake. The earthquake-specific criterion and the final classification of each data point are shown on Figure D.3.

REFERENCES

- Hastie T. J. (1992). *Generalized Additive Models*, Chapter 7 of Statistical Models. In: J. M. Chambers, T. J. Hastie (eds.), Wadsworth & Brooks/Cole.
- Koenker R. (2005). *Quantile Regression*, Cambridge University Press: Cambridge, 349 pp., doi: 10.1017/CBO9780511754098.
- Petersen M. D., Dawson T. E., Chen R., Cao T., Wills C. J., Schwartz D. P., Frankel A. D. (2011). Fault displacement hazard for strike-slip faults. *Bulletin of the Seismological Society of America*, **101**(2), 805–825, doi: 10.1785/0120100035.

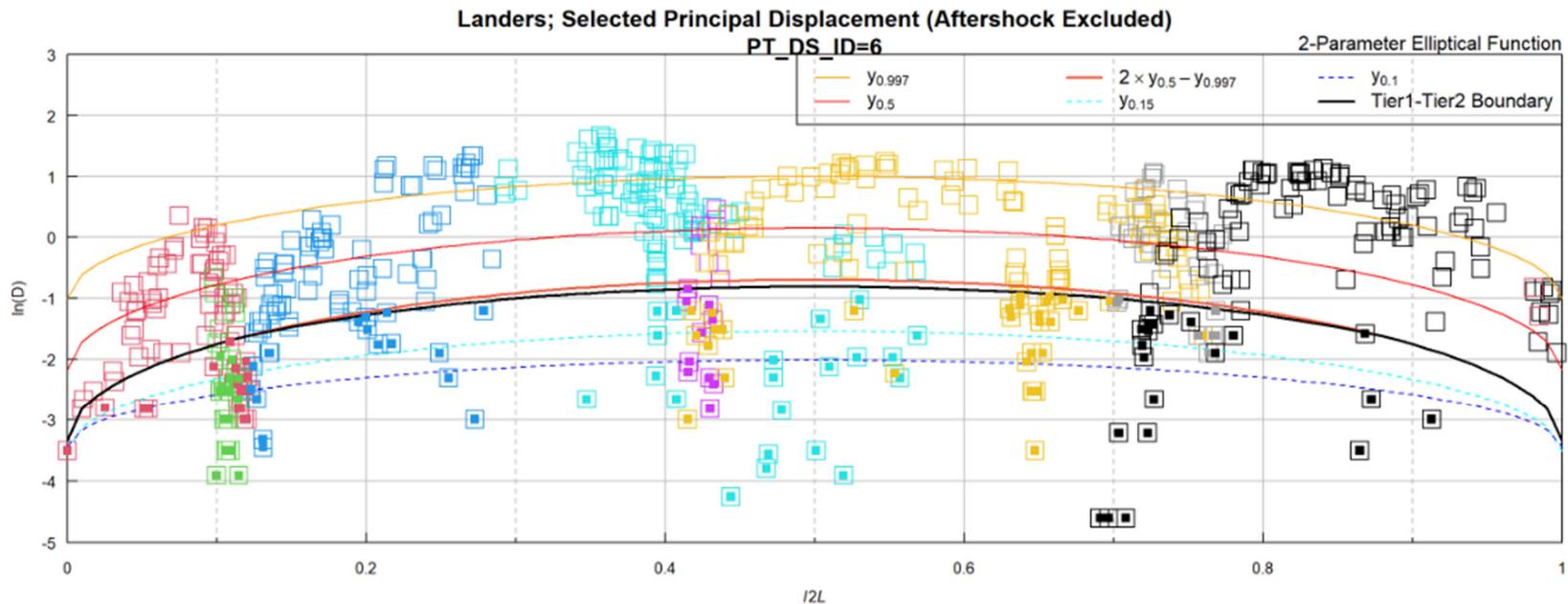


Figure D.1. An example of tier classification of the Landers principal net displacement (D , in units of meters) from the preferred data source PT_DS_ID = 6, using the 2-parameter ellipse functional form. The estimated 0.997 and 0.5 quantiles are shown as the orange and red curves, respectively. The estimated 0.003 quantile of the hypothetical normal distribution is shown as the thick red curve, while the lower-bound of tier-1 displacements is shown as the thick black curve. The identified tier-2 displacements are marked by a small solid square inside an open square. Displacements from the same rupture segment are marked by the same color. For comparisons, the 0.15 and the 0.1 quantiles are shown as the cyan and the blue dashed lines, respectively.

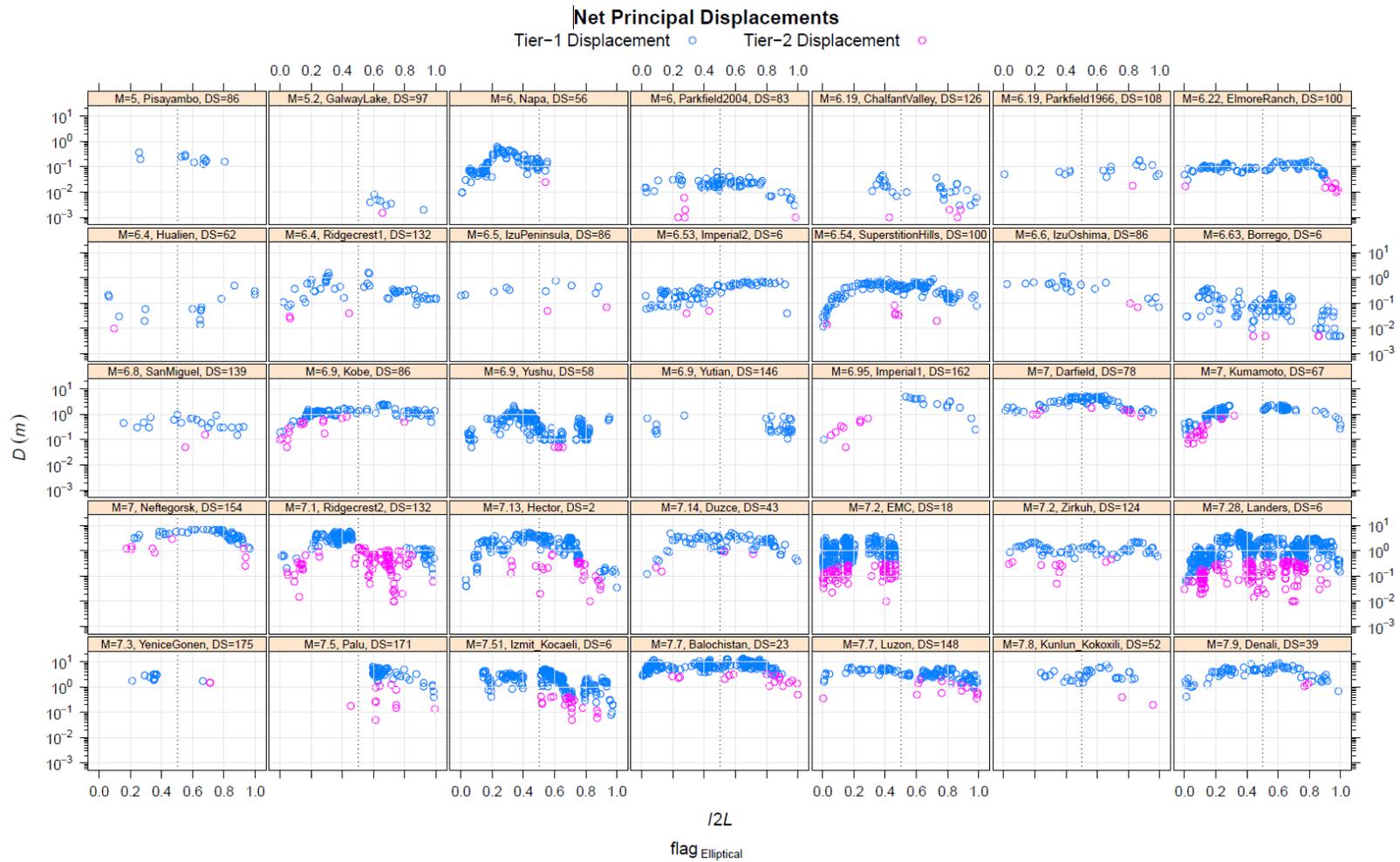


Figure D.2. Tier classification of principal net displacement (D) obtained using each of the eight alternative functional form of $f_{|2L}(|2L)$: (A) Elliptical. (B) Elliptical plus Linear. (C) Quadratic. (D) Cubic. (E) Modified Beta. (F) B-Spline, degrees of freedom (df) = 3. (G) B-Spline df = 5. (H) B-Spline df = 7. Earthquake name, magnitude (M), and preferred FDHI data source identification (DS) are indicated inside each panel title strip.

Note: This page is for Figure D.2(A). Figures D.2(B) through D.2(H) are in the next 7 pages

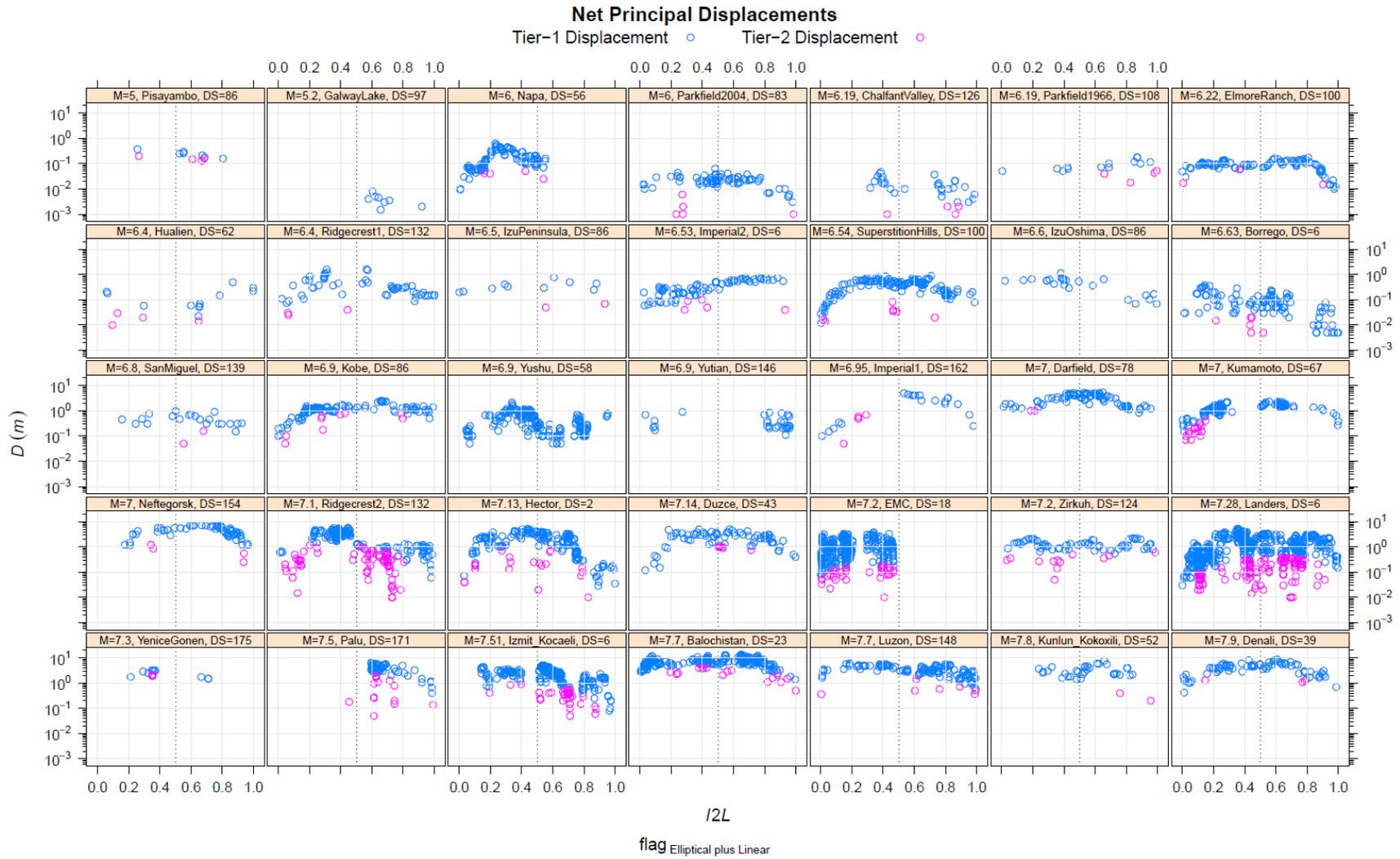


Figure D.2. Tier classification results obtained using each of the eight alternative functional form of $f_{l_2L}(l_2L)$: (B) Elliptical plus linear.

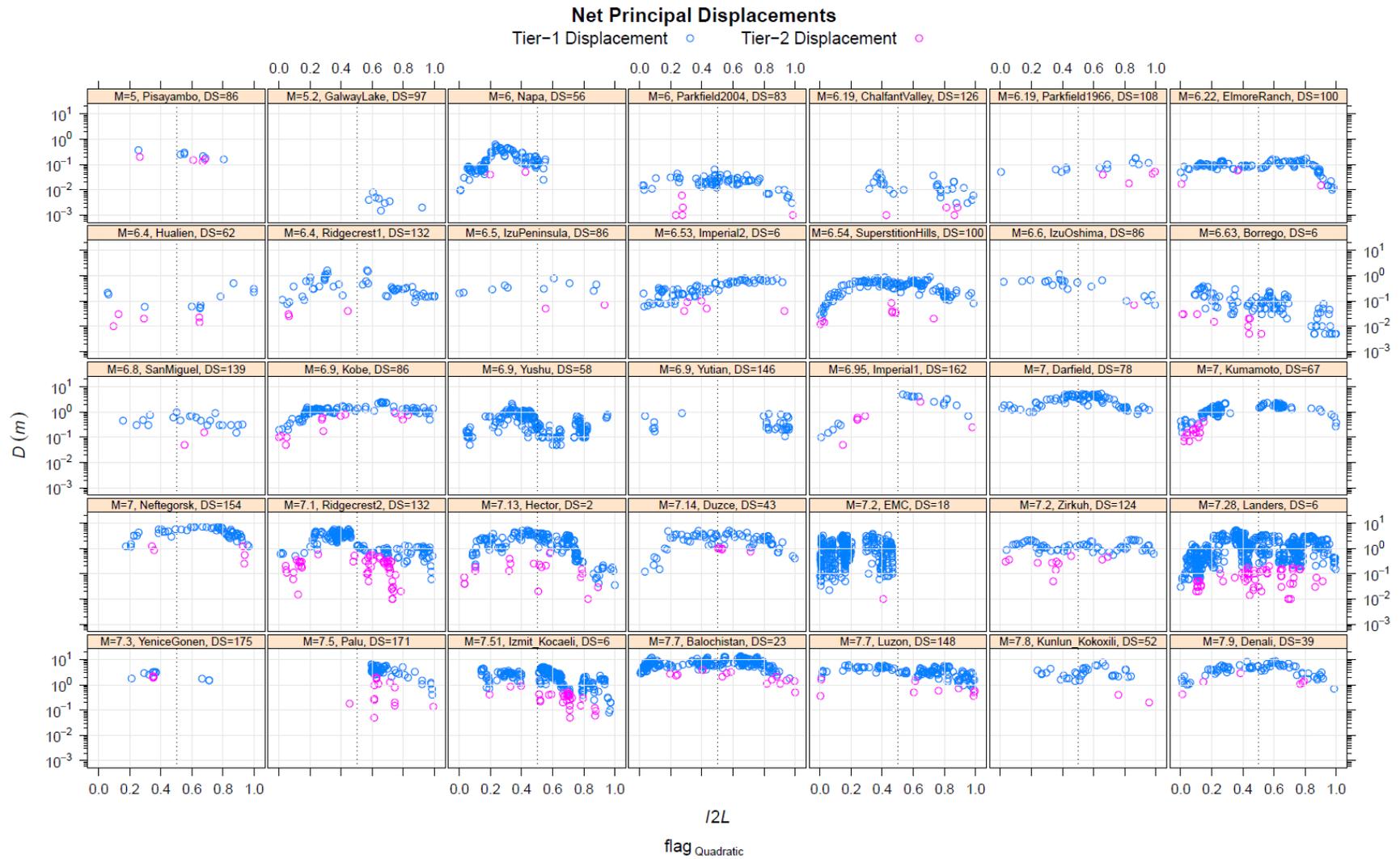


Figure D.2. Tier classification results obtained using each of the eight alternative functional form of $f_{l2L}(l/2L)$: (C) Quadratic.

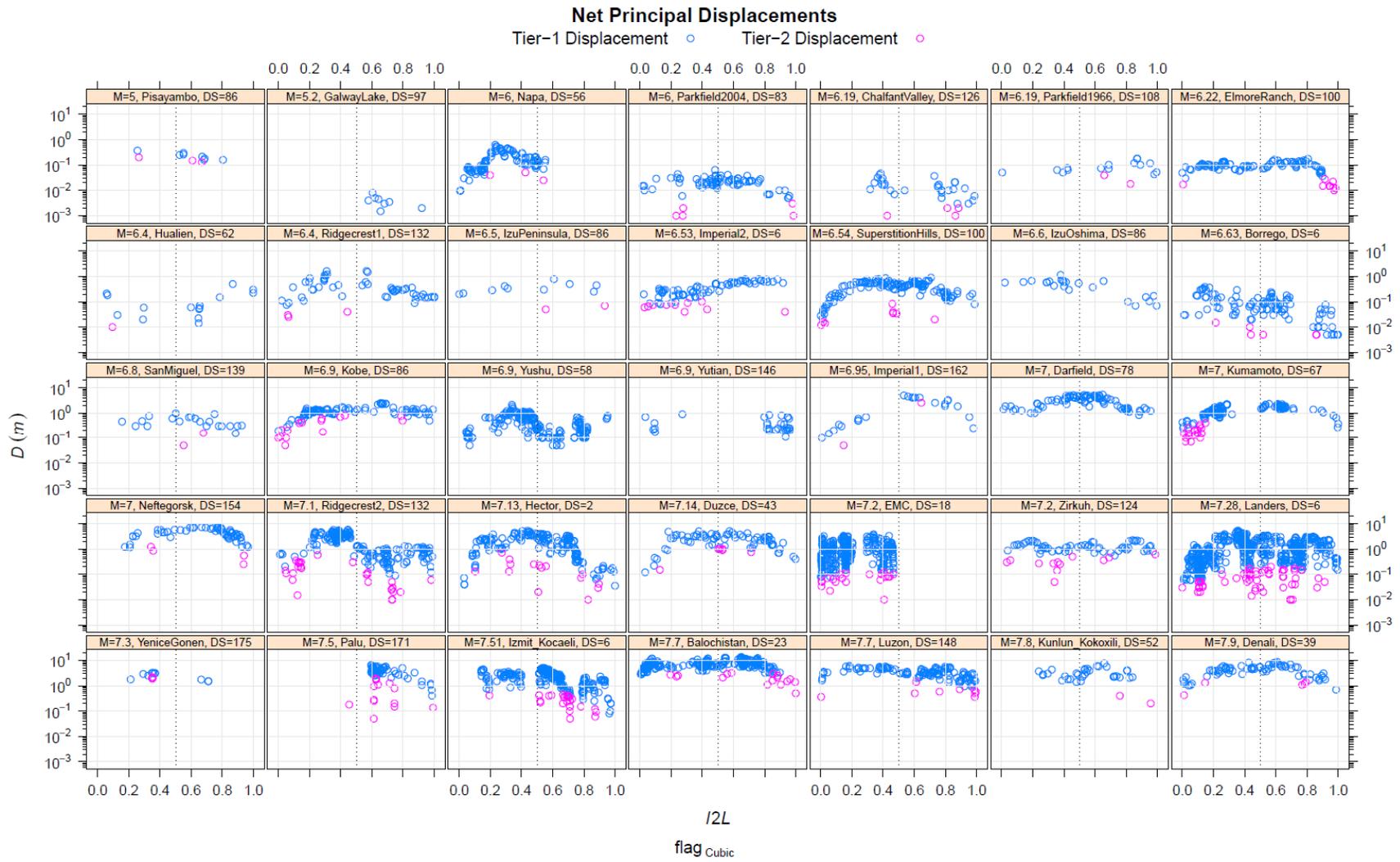


Figure D.2. Tier classification results obtained using each of the eight alternative functional form of $f_{l2L}(l2L)$: (D) Cubic.

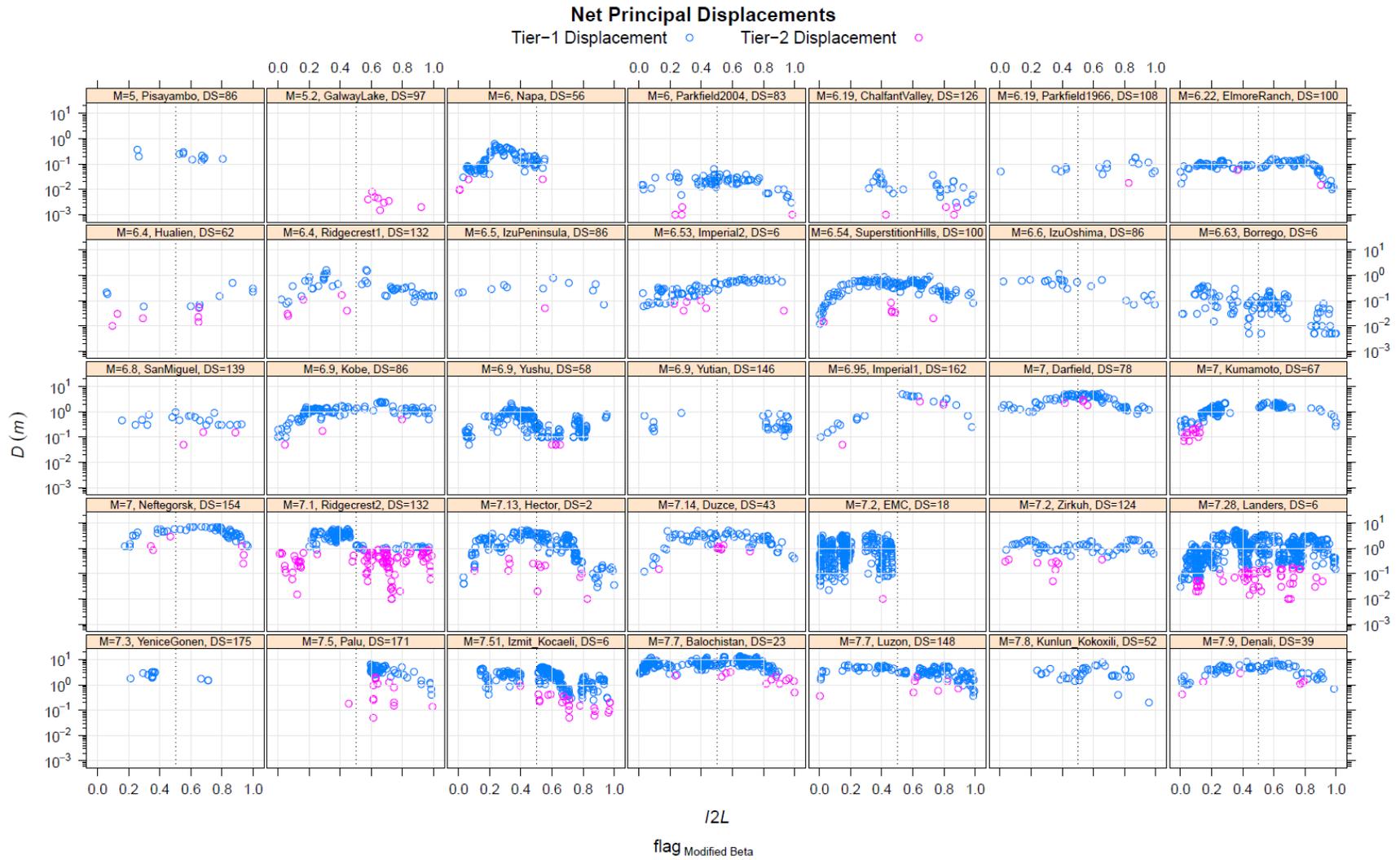


Figure D.2. Tier classification results obtained using each of the eight alternative functional form of $f_{l/2L}(l/2L)$: (E) Modified Beta.

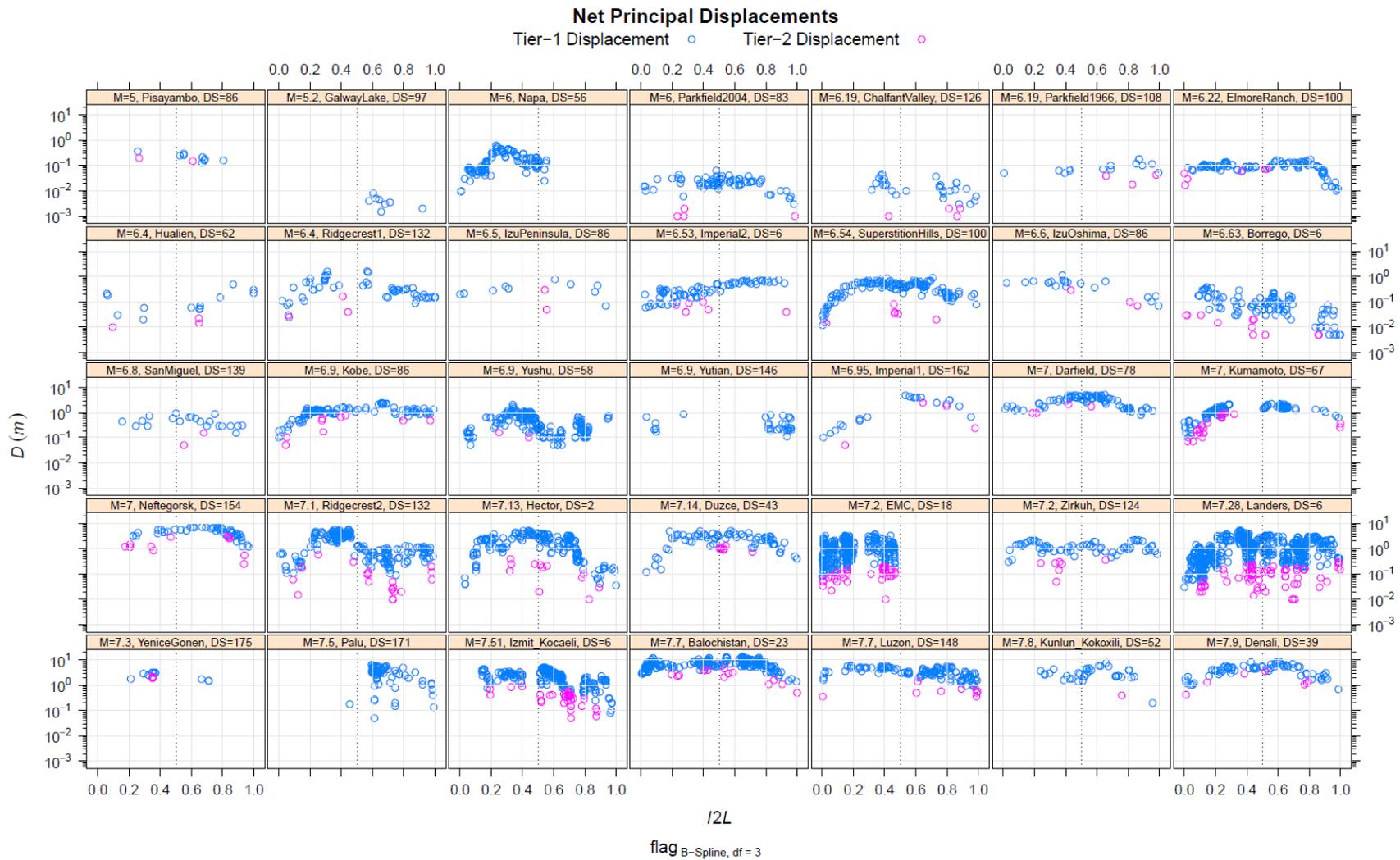


Figure D.2. Tier classification results obtained using each of the eight alternative functional form of $f_{I_{2L}}(I_{2L})$. (F) B-Spline, $df = 3$.

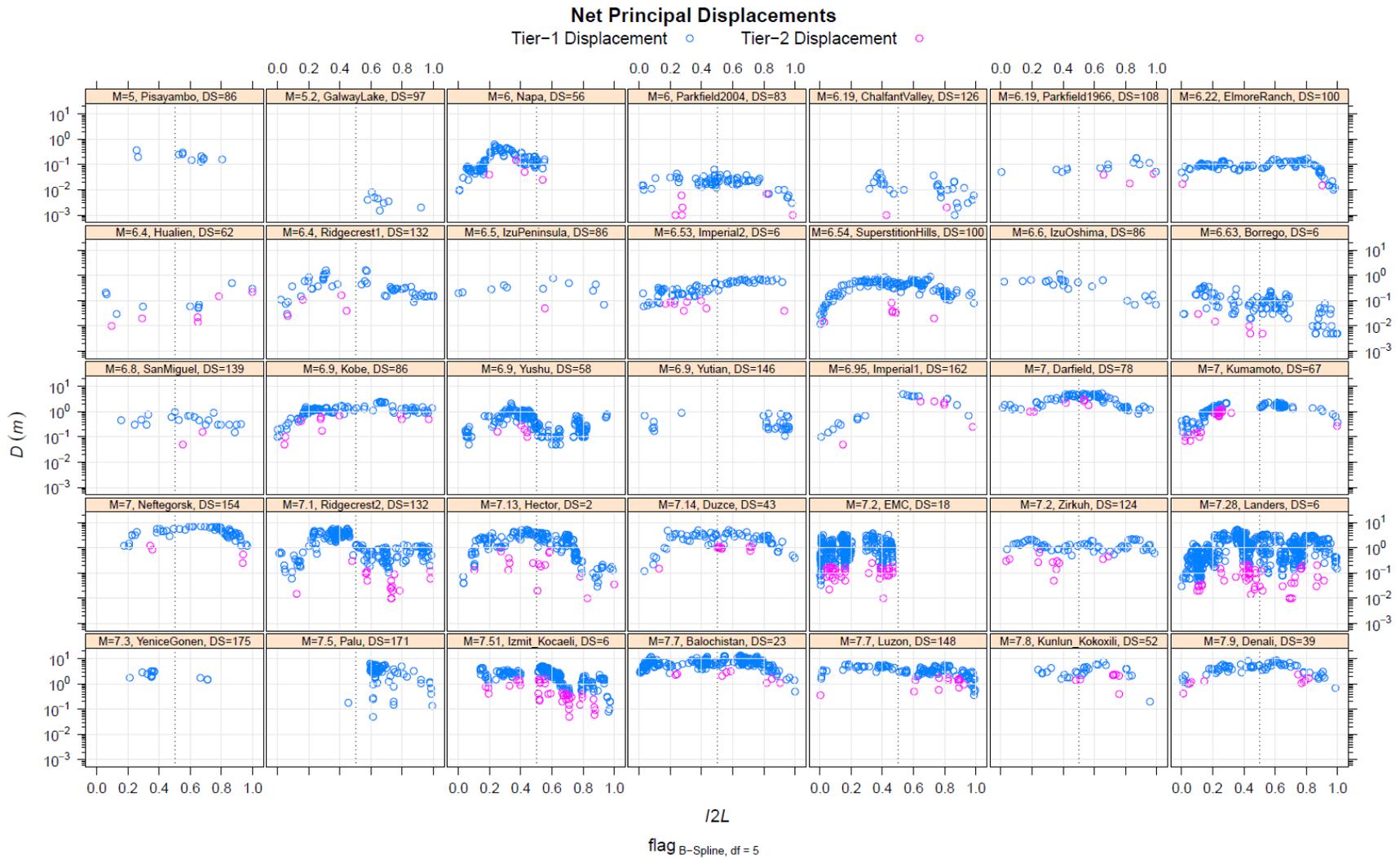


Figure D.2. Tier classification results obtained using each of the eight alternative functional form of $f_{l2L}(l2L)$: (G) B-Spline, $df = 5$.

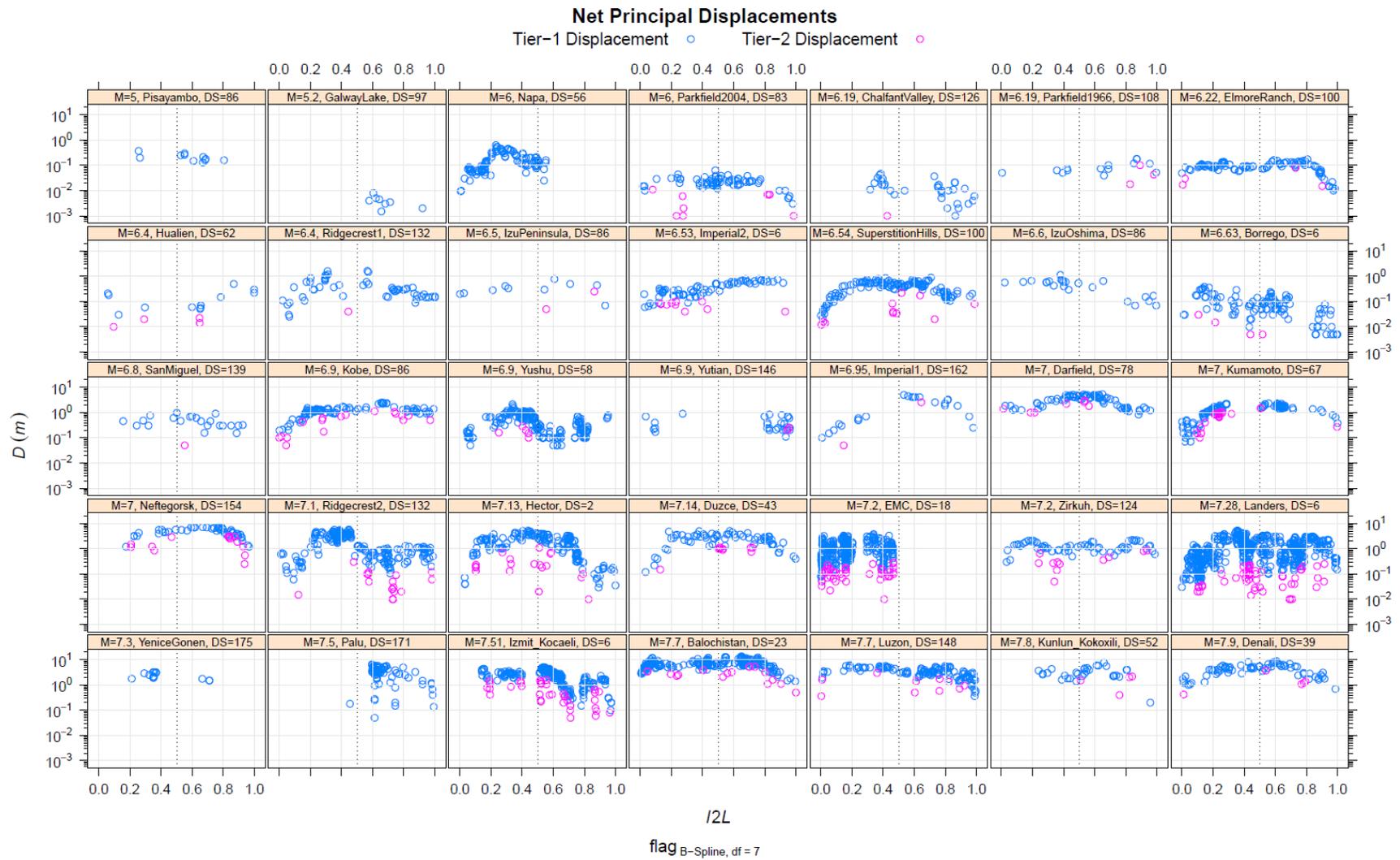


Figure D.2. Tier classification results obtained using each of the eight alternative functional form of $f_{l2L}(l2L)$: (H) B-Spline, $df = 7$.

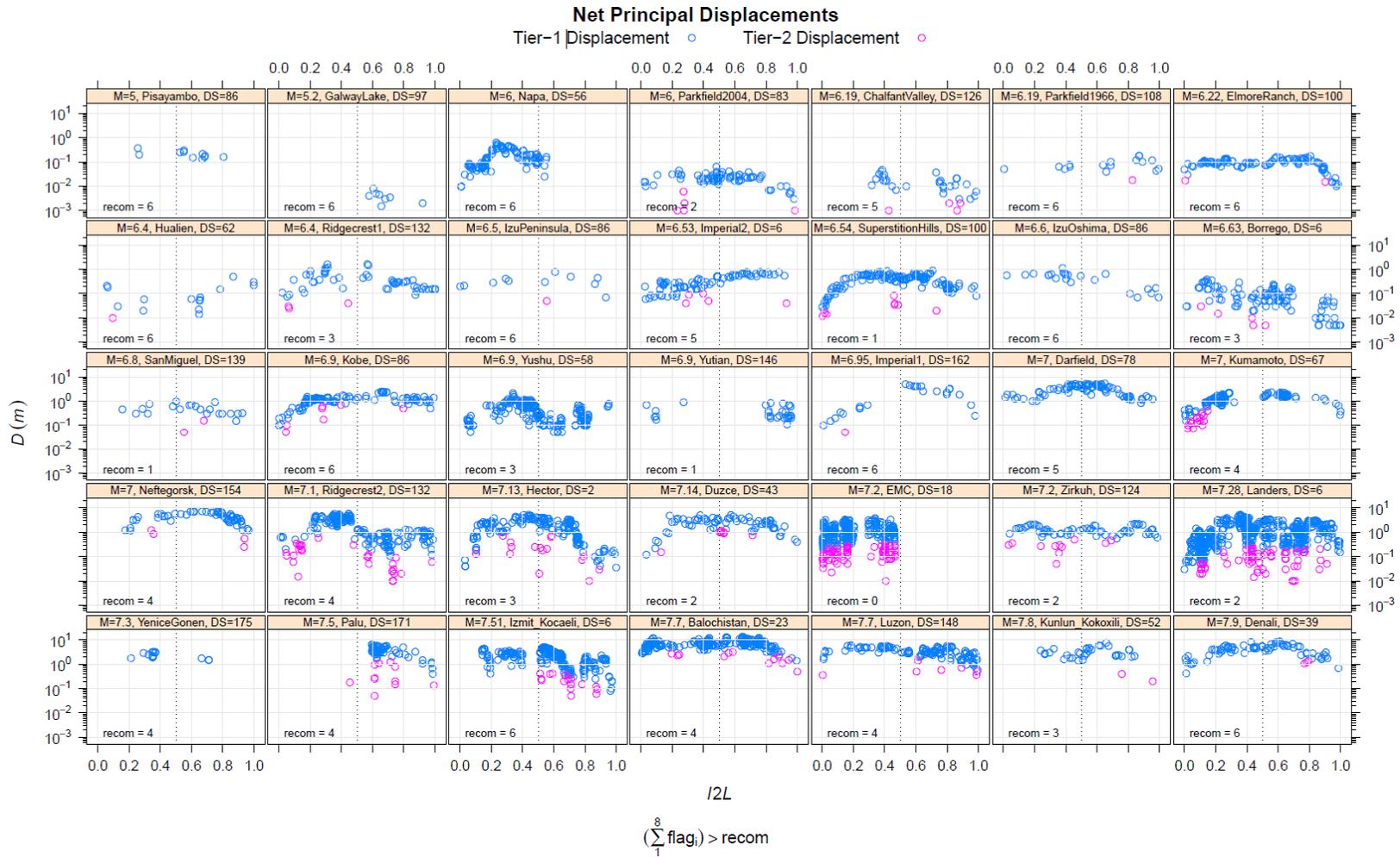


Figure D.3. Single final tier classification of a data point. Earthquake name, magnitude (M), and preferred FDHI data source identification (DS) are indicated inside the panel title strip. Indicator variable $flag_i$ is equal to 1 if the data point of interest is classified as tier 2 by the i -th functional form, otherwise $flag_i = 0$. The sum $(\sum_{i=1}^8 flag_i)$ is the number of times a data point is classified as tier 2. A data point is assigned a final classification of tier-2 if $(\sum_{i=1}^8 flag_i) > recom$; that is, it is classified as tier 2 more than $recom$ times. The criterion $recom$ is earthquake specific and given inside each panel.

ELECTRONIC SUPPLEMENTS

We provide two electronic supplements at <https://www.conservation.ca.gov/cgs/pfdha>:

1. Selected measurement data used in model development in comma-separated value file format
2. ArcGIS shapefiles of the least-cost paths for all strike-slip events in the Fault Displacement Hazard Initiative (FDHI) database
3. Least-cost path analysis – ArcGIS model and a step-by-step procedural guide