5
CHAPTER 5
SSC MODEL: OVERVIEW AND METHODOLOGY

5.1 Overview of Spatial and Temporal Models

The heart of any SSC model for PSHA is a description of the future spatial and temporal distribution of earthquakes. The earliest PSHA models identified seismic source zones within which the spatial distribution of earthquakes was assumed to be uniform. The temporal distribution of earthquakes was assumed to follow a Poisson process, and the sizes of earthquakes within the zones were assumed to follow an exponential distribution. In more tectonically active areas such as the WUS, the recognition of active faults allowed the spatial models to evolve such that they were able to include the fault sources as well as the background source zones within which they lie. Further, the development of fault-specific paleoseismic data regarding the timing, amount of slip per earthquake, and geologic slip rate provided a new class of temporal models that could use this information directly in the PSHA.

In the meantime, the evolution of spatial and temporal models within stable continental regions (SCRs) such as the CEUS has taken place differently and more slowly. Despite continued geologic investigations, the search for the causative faults giving rise to observed seismicity has yielded very few cases where a definitive argument can be made for having identified an active fault in the sense that this is commonly identified in the WUS. With notable exceptions such as the Meers fault and the Cheraw fault, the best-defined cases of localized seismicity and deformation occur in the New Madrid region, where individual fault-like sources can be called out and characterized. Even in locales such as Charleston, where geologic and geophysical studies have continued over the decades, uncertainties persist regarding the unique association of the observed historical seismicity and the one or more causative faults that are responsible.

5.1.1 Spatial Model Considerations

Continued analysis of the historical seismicity record and network monitoring by regional and local seismic networks has led to acceptance within the community that the general spatial patterns of observed small-to-moderate magnitude earthquakes provide predictive information about the spatial distribution of future large-magnitude earthquakes. The analyses leading to this conclusion have focused on whether the observed patterns of earthquakes have varied through time; therefore, in effect, this is an assessment of whether small- to moderate-magnitude earthquakes have been relatively stationary through time. Studies such as those by Kafka (2007, 2009) suggest that this has generally been the case and that we can with some level of confidence use the spatial pattern of smaller earthquakes to predict the future pattern of smaller earthquakes.

However, the available data on larger-magnitude earthquakes and their relationship to the spatial distribution of smaller earthquakes based on the observed record is quite limited and not sufficient to allow an empirical spatial model to arrive at such a conclusion with confidence. We have some cases like Charlevoix and New Madrid where a spatial concentration of smaller-magnitude earthquakes is occurring generally within the region of larger historical earthquakes. However, we also have counter-examples like the Eastern Tennessee seismic zone, which is defined by a highly concentrated zone of small to moderate earthquakes without any larger events in the historical record. Likewise, the Charleston region is associated with a pattern of observed seismicity that is not particularly remarkable for drawing attention to the location of the 1886 earthquake.

Perhaps the most dramatic finding in recent years is paleoseismic evidence for the occurrence of prehistoric large-magnitude earthquakes in the CEUS. For obvious reasons, the earliest paleoseismic studies occurred in the locales of large historical earthquakes (e.g., New Madrid, Charleston), and subsequent studies moved to other regions (e.g., Wabash Valley, Charlevoix, Eastern Tennessee). With a few exceptions, the paleoseismic information developed at these locations appears to be related to secondary shaking effects such as liquefaction rather than observed fault displacements. As a result, uncertainties exist about both the locations and sizes of the earthquakes. Detailed studies of historical and instrumental earthquakes, and liquefaction phenomena associated with them, coupled with field and laboratory studies of geotechnical properties, are leading to a stronger technical basis for placing limits on the locations of paleoearthquakes interpreted by the distribution of liquefaction phenomena and for defining their magnitudes. In some cases, the paleoseismic evidence for repeated large-magnitude earthquakes is so compelling that the TI Team has concluded that a RLME source must be included in the SSC model. In other cases, there is suggestive evidence of potential liquefaction earthquakes, such as the postulated Alabama-Louisiana-Mississippi (ALM) zone. In that situation, consideration is given to possible evidence for a prehistoric earthquake, but inclusion as an RLME source is not warranted (see Section 7.3.9 for a discussion). The establishment of several RLME sources as spatial components of the CEUS SSC model is an important advancement over previous studies and reflects the evolution taking place within the technical community.

Notwithstanding the locations of RLME sources, the spatial distribution of distributed seismicity sources has advanced in PSHA based largely on the assumption of spatial stationarity, as discussed in Section 5.3.2.1 of this report. For example, the EPRI-SOG project and the U.S. National Seismic Hazard Mapping Project (Petersen et al., 2008) use different approaches to "smooth" observed seismicity to provide a map that expresses the future spatial pattern of seismicity. In reality, the *a*-values, or rates per unit area, and *b*-values are subject to the smoothing operation, and various calculation approaches are carried out. Because of the support in the community for these approaches to representing the future spatial distribution of seismicity, they have been explored thoroughly as part of the CEUS SSC Project. Various adaptive kernels were evaluated along with the penalized maximum-likelihood approach that is a refinement of the approach used in the EPRI-SOG project. The refined approach subdivides the region into one-quarter-degree cells and specifies the degree of "communication" between adjacent cells. As discussed in Section 5.3.1, the approach was selected because, although it is based on the same conceptual model of spatial stationarity as the kernel approach, it offers the

ability to allow for spatial variability in both *a*- and *b*-values, and it was found to provide comparable results to the adaptive kernel when a constant *b*-value was assumed.

5.1.2 Considerations Regarding Temporal Models

For purposes of SSC model assessments, the "temporal" issue concerns both the sizes of future earthquakes as well as their frequency of occurrence (average rate and aperiodicity). With time and study the earthquake research community in the WUS has developed information that suggests that the assumption of an exponential distribution of magnitudes may not be appropriate for individual faults. Poissonian models are memoryless and do not incorporate information about the timing or size of prior earthquakes. Studies of fault displacements, including the timing and displacement per event, and physical notions of strain accumulation and release provide information for modeling where a source currently resides in the seismic cycle. Such "real-time" and "renewal" models have found their way into some SSC model assessments where the dominant sources are highly active plate-boundary faults.

In contrast to plate-boundary SSC models, the fundamental information source for the assessment of temporal models in the CEUS is the catalog of historical and instrumental earthquakes. Though uneven in coverage and incomplete back through time, the catalog is the cornerstone for assessing recurrence rates for most seismic sources in the CEUS. Thus, significant effort was devoted to developing an earthquake catalog for the CEUS SSC Project with moment magnitudes for all earthquakes and for analyzing its completeness as a function of time and location. Beyond the catalog, in the CEUS, the modeling of temporal behavior has benefited most from paleoseismic investigations at the RLME sources. An ever-increasing database of paleoliquefaction studies, which have been compiled as part of the CEUS SSC Project (Appendix E), provides information on the approximate sizes and timing of large-magnitude events. In effect, these studies extend the historical observations back a few thousand years. Despite their uncertainties, the studies provide clues about the nature of the repeated pattern of larger earthquakes through time at individual seismic sources. A number of studies conducted within SCRs including the CEUS have concluded that large earthquakes are not evenly spaced in time, but rather occur as "clusters" of earthquakes that are separated from other clusters by long "intercluster" periods of quiescence.

For each RLME source, the available data were evaluated in order for the TI Team to assess whether temporally clustered behavior has been observed and, if so, whether the source is currently within or outside of a temporal cluster. In either case, the temporal behavior is modeled as a Poisson process. The TI Team further evaluated the available data to assess whether the application of a renewal model would be appropriate for the RLME sources. The physical underpinning of a renewal model is a quasi steady state loading process applied to a fault. These models have been given relatively high weight in characterizing earthquake occurrences on plate boundary faults, even though there may not be a well-developed paleoseismic history (e.g., Working Group, 2003). Two of the RLME sources defined by the TI team have reasonably well-developed paleoseismic histories: New Madrid and Charleston. However, the physical process of steady state loading being applied to these sources is more problematic. In addition, the specific causative faults associated with the New Madrid RLME sources are only moderately understood and the fault or faults associated with the Charleston RLME source are poorly understood.

Therefore, giving due consideration to the larger technical community's views regarding recurrence models for these RLME sources, the TI team concluded that applying the renewal model to these two RLME sources should have some weight, although it is relatively low. For all of the other RLME sources, the available data were judged not to be sufficient to allow for the application of such models with any confidence. The methodology for characterizing RLME recurrence is given in Section 5.3.3 and the specific data and recurrence assessments related to the Charleston and New Madrid RLME sources are given in Sections 6.1.2 and 6.1.5, respectively.

For distributed seismic source zones, the earthquake catalog provides the fundamental basis for assessing recurrence and, as described previously, for describing its future spatial distribution. Consistent with common practice in the assessment of seismicity within large seismic sources, the observed seismicity is assessed to follow a truncated exponential magnitude distribution (e.g. Cornell and Van Marke, 1969) and Poissonian behavior.

5.1.3 Perspective on CEUS SSC Models

At the time of the preparation of this report, there were a number of promising research areas pertaining to spatial and temporal models that had not reached the level where they could be included in the SSC model with confidence. Examples include models proposed as part of the ongoing debate on "characteristic" earthquakes or, simply stated, the tendency for a specific fault to generate earthquake magnitudes that are not exponentially distributed in magnitude. The TI Team reviewed this continuing dialogue within the community, but concluded that the models were not applicable to the kinds of sources being characterized in the CEUS. Models of fault-specific behavior were not judged to be applicable in the CEUS unless and until seismogenic faults were identified and data regarding their behavior was gathered. Rather, the intent of the RLME sources is to model only the recurrence of repeated large earthquakes in a specific location. As such, the RLME sources represent a source of seismic hazard that is in addition to the hazard resulting from the occurrence of distributed seismicity modeled by the distributed seismicity sources. The addition of the RLME recurrence with that derived from observed seismicity within the distributed seismicity sources allows for a departure from the general truncated exponential recurrence behavior in specific locations. This simple approach to incorporating the data available in the CEUS is described in Sections 5.3.3.4 and 6.1.

Another area of ongoing research with potential implications for recurrence behavior relates to geodetic strain rate measurements. No doubt, the systematic gathering of geodetic data throughout the CEUS for purposes of identifying and analyzing tectonic strain is crucial for the understanding of contemporary deformation processes. Its importance is acknowledged and its promise in the decades ahead is undeniable. However, in 2010, the CEUS technical community's understanding of the meaning of a short-term (few decades) strain signal relative to seismic source characteristics is embryonic. The spatial distribution and temporal stability of such strains is not known, and there is no clear method for translating these strain signals into spatial and temporal SSC models. One need only witness the ongoing debates in the WUS on the discrepancies between long-term geologic slip rates and short-term geodetic strain rates to lose confidence that, in the complete absence of geologic slip rates, the geodetic results can be used.

As a corollary, the types of analyses carried out in the WUS to "test" the reasonableness of a seismic source model are generally not applicable in the CEUS. For example, the integrated moment rate derived for a region from the summed fault slip rates (derived from geologic studies) cannot be developed and, as a result, cannot be compared to the observed moment rate for the same region based on the seismicity record. Likewise, given the generally sparse geodetic networks and short period of observation, the uncertainties in attempting to compare the recurrence rates for seismic sources derived from seismicity with those developed from geodetic data are unacceptably large. It is acknowledged in this report that local well-studied areas such as the central New Madrid region may eventually provide clues or even constraints on the temporal behavior of seismic sources. Research in these areas is encouraged, but the current data do not allow for direct incorporation into the CEUS SSC model.

5.2 Maximum Earthquake Magnitude Assessment

This section describes the methodology that was used in the CEUS SSC Project to assess Mmax for all seismic sources. Estimating Mmax in SCRs such as the CEUS is highly uncertain despite considerable interest and effort by the scientific community over the past few decades. Mmax is defined as the upper truncation point of the earthquake recurrence curve for individual seismic sources, and the typically broad distribution for any given source reflects considerable epistemic uncertainty. Because the maximum magnitude for any given seismic source in the CEUS occurs rarely relative to the period of observation, the use of the historical seismicity record provides important but limited constraints on the magnitude of the maximum event. At annual frequencies of interest for conventional buildings (approximately 10^{-2} to 10^{-3}/yr), the hazard results are usually not particularly sensitive to the maximum magnitude. However, at annual frequencies of interest for nuclear facilities (10^{-3} to 10^{-7}), the influence of Mmax can be very significant. For this reason, focus on the Mmax issue has been driven by PSHAs conducted for nuclear facilities.

In view of the importance of Mmax revealed by the EPRI-SOG project (EPRI, 1988), a major study sponsored by EPRI was carried out in the early 1990s to attempt to quantify the knowledge and uncertainties in Mmax for purposes of PSHA (Johnston et al., 1994). The approach advocated in the EPRI Mmax study is based on analogies between the CEUS and other SCR areas. To develop the analogies, two key activities needed to be conducted: (1) development of a worldwide seismicity catalog of moderate to large earthquakes with systematic moment magnitudes for all events, and (2) identification and characterization of the regions worldwide that can be considered analogous tectonically to the CEUS. The results of those efforts were the first of their kind and led to considerable advances in the research (e.g., Johnston, 1996a).

The EPRI Mmax study also focused on the use of a Bayesian procedure for estimating Mmax, in which the prior distribution is based on the earthquake magnitudes of events that occurred worldwide within tectonically analogous regions. At the time the study was conducted, there was considerable focus on the occurrence of large earthquakes within formerly rifted or extended crust. For example, the 1811-1812 New Madrid earthquakes were interpreted to have occurred within the Reelfoot rift, which exhibits evidence of multiple episodes of extension but is currently within a compressive stress regime. Despite the anecdotal evidence for the occurrence of the largest SCR earthquakes within extended crust, the statistical analyses in Johnston et al. (1994) did not provide strong evidence that any particular tectonic characteristic—or

combination of characteristics—is correlated with maximum earthquake size. The best correlations at the time suggested that sites of major Paleozoic and younger extension appeared to have the potential for a larger Mmax (i.e., a higher mean of the prior distribution), but the analyses also indicated that the standard deviation of the prior distribution was larger.

From the beginning of the CEUS SSC Project, the TI Team pursued the refinement and application of the Bayesian Mmax approach (Johnston et al., 1994). An evaluation of alternative approaches to assessing Mmax in the CEUS (Wheeler, 2009) concluded that the Bayesian Mmax approach, as well as other approaches that were based on tectonic analogues, was worthy of pursuit within the technical community. The Bayesian Mmax approach has the advantage that it provides a quantitative and repeatable process for assessing Mmax. The approach is described below in Section 5.2.1.1 along with the updates to the Johnston et al. (1994) characterization of Mmax for SCRs. The TI Team also explored alternative approaches for the assessment of Mmax that provide quantitative and repeatable results, and the team identified the approach developed by Kijko(2004) as a viable alternative. In his evaluation of statistical approaches for estimating Mmax, Wheeler (2004) noted that problems of small samples and rare large earthquakes continue to hinder testing of statistical models in SCRs. Nevertheless, the Kijko (2004) approach requires fewer assumptions than the Bayesian approach in that it uses only the observed earthquake statistics for the source, but this is offset by the need for a relatively larger data sample in order to get meaningful results. The Kijko (2004) approach is described in Section 5.2.1.2.

5.2.1 Approaches to Mmax Estimation in the CEUS

The CEUS SSC model incorporates two types of seismic sources: RLME sources and distributed-seismicity source zones of various types. The RLME sources are characterized by their sizes and recurrence rates derived primarily from the paleoseismic record and, despite variations in the seismic source zone configurations, are assumed to always exist in the model. RLME sources include those in the New Madrid area, Wabash Valley, Charleston, Charlevoix, and Meers fault, among others. Because of the independent constraints on earthquake size, those constraints are used to estimate the magnitudes of RLMEs, and the assessment of Mmax discussed here is not applicable to RLME sources. Rather, this discussion is applicable to the distributed-seismicity source zones, which exist in different forms depending on whether the Mmax zones or the seismotectonic zones branches of the master logic tree are being followed.

The assessment of Mmax for distributed seismicity sources (source zones) in the CEUS SSC Project has been designed to incorporate the uncertainties in both the conceptual models and the parameter values associated with each model. Based on the review and evaluation of the Bayesian method and consideration of other methods that exist within the larger technical community, it was decided that the assessment would need to include alternative conceptual models for Mmax. In particular, it was decided that alternative prior distributions should be considered for the Bayesian approach, and that an alternative to the Bayesian approach should be considered.

The TI Team decided to use two alternative approaches as a basis for estimating the Mmax distributions for distributed seismicity sources: (1) the Bayesian procedure, which uses the prior

distributions from an update and evaluation of the global SCR data; and (2) the Kijko (2004) procedure, which uses the observed seismicity within a region to provide a direct (or posterior) assessment of Mmax. The Bayesian approach is representative of a category of approaches that rely on drawing analogies to tectonically comparable regions in order to estimate the Mmax for the source of interest. These approaches are based on the ergodic assumption that one can substitute time (the short period of observation) for space (other SCRs). The Kijko approach is representative of an alternative category of models that rely on the observations of seismicity entirely within the zone of interest. Because the occurrence of Mmax is typically rare relative to the period of observation, these approaches assume a particular frequency distribution of earthquake sizes and, to be able to apply them with confidence to larger or very active regions, rely on significant numbers of observed earthquakes to provide stable estimates.

Both the Bayesian and Kijko approaches have their pros and cons. They both have the positive attribute that they are repeatable given the same data and that they can readily be updated given new information. The Bayesian approach is arguably more stable because of the use of a prior distribution that, even in the absence of a significant number of earthquakes in the zone of interest, can still provide a result. However, the prior distributions for Mmax are developed based on analogies to other "tectonically comparable" regions, which can be a source of uncertainty, and based on evaluation of those regions relative to a highly uncertain set of characteristics that are postulated to be important to Mmax. The advantage of the Kijko approach is that it does not require the identification of analogue regions or assessments of the characteristics of those regions. However, as applied the approach relies on the assumption that the distribution of earthquake magnitudes follows a doubly truncated exponential distribution. Moreover, the approach does not provide stable results when the number of observed earthquakes is low.

From the standpoint of their specific use in the CEUS SSC model, there are other considerations in evaluating the two approaches. The key difference between the Kijko approach and the Bayesian approach is that the Kijko approach does not have a prior distribution; it uses only the earthquakes within the source of interest. The prior distribution is based on analogies to other parts of the world and the assumption that those events are applicable to the estimation of maximum magnitudes within the source of interest. A potential problem is that the global SCR database includes earthquakes from RLME sources (e.g., New Madrid, Charleston, and perhaps others worldwide that we suspect might have RLMEs). In the CEUS SSC model, however, the Bayesian approach is being applied to non-RLME sources, so we are possibly overestimating Mmax by using RLMEs globally to develop our prior distributions. The Kijko approach has some appeal because, by definition, it uses only data from the CEUS and only the earthquake catalog that does not include RLMEs. Care must be taken to avoid instability due to small numbers of earthquakes, but this can be accomplished by examining the statistics of the calculations, as provided for in the Kijko approach, and by giving the approach less weight when numbers of events are too low for stable estimates. Also, the seismic sources of interest in the CEUS SSC model tend to be larger regions with a significant number of earthquakes, thus mitigating the likelihood of instability.

The application of the Bayesian and Kijko approaches to assessing Mmax for the CEUS SSC Project is described below in Sections 5.2.1.1 and 5.2.1.2. The approach to weighting the alternative approaches for particular seismic sources is described in Section 5.2.1.3.

5.2.1.1 Bayesian Mmax Approach

The Bayesian Mmax approach is illustrated on Figure 5.2.1-1. The approach is based on estimating Mmax by drawing analogies to regions that are comparable to the CEUS (i.e., SCRs) and using that information to establish *prior* distributions on Mmax for distributed seismicity sources (source zones) in the CEUS. Part (a) of Figure 5.2.1-1 shows the *prior* distribution developed by Johnston et al. (1994) for extended continental crust. The prior distribution is then updated with source-specific information about the number and sizes of observed earthquakes. The update is expressed as a *likelihood* function, part (b) of Figure 5.2.1-1, which is zero for all values below the maximum observed magnitude, has a peak near the maximum observed, and then becomes essentially constant at larger magnitudes. The prior distribution is then convolved with the likelihood function to arrive at a *posterior* distribution of Mmax, shown in part (c) of the figure. The continuous posterior distribution is then represented by a discrete distribution for use in hazard analysis, as shown in part (d).

The likelihood function used to update the prior is derived from the earthquake recurrence model applied to the source region. The model used in this study is an exponential distribution for the sizes of earthquakes occurring with a source region. This is the standard model applied to regions where multiple individual sources contribute to the occurrence of earthquakes. The likelihood function for possible values of Mmax, m^u, based on the exponential model is (e.g., Johnston, 1994) is

$$L[m^u] = \begin{cases} 0 & \text{for } m^u < m_{\max-obs} \\ \left[1 - \exp\left\{-b\ln(10)(m^u - m_0)\right\}\right]^{-N} & \text{for } m^u \geq m_{\max-obs} \end{cases} \quad (5.2.1\text{-}1)$$

where b is the Gutenberg-Richter b-value, N is the number of recorded earthquakes with magnitudes equal to or larger than a minimum value m_0, and $m_{\max-obs}$ is the largest recorded earthquake. The likelihood function has two effects on the prior. The first is to truncate the lower tail of the prior distribution at the value of the maximum observed earthquake. The shape of the likelihood function is controlled by the value of N, the number of earthquakes between m_0 and $m_{\max-obs}$. In most cases, the value of N is small such that the likelihood function is relatively flat with a small peak at m^u equal to $m_{\max-obs}$, as shown in part (b) of Figure 5.2.1-1. As a result, the primary effect is to produce a posterior distribution with a similar shape to the prior, as shown in part (c) of Figure 5.2.1-1. When N becomes large, the likelihood function becomes very peaked at $m_{\max-obs}$ and has a major effect on the prior distribution. Figure 5.2.1-2 repeats the example of Figure 5.2.1-1 with the value of N increased from 2 to 10. In this case, the posterior distribution also becomes very peaked near $m_{\max-obs}$. It is important to note that the influence of size of N is relative to the magnitude range m_0 to $m_{\max-obs}$. Because of the exponential distribution of magnitude sizes, increasing N by reducing m_0 for a given value of $m_{\max-obs}$ has no effect on the

likelihood function as long as the increase in N is consistent with the b-value of the source region.

Because typical applications involve small values of N, the shape of the posterior Mmax distribution is similar to that of the prior. Thus, specification of the prior distribution is of primary importance in defining the Mmax distribution. The Johnston et al. (1994) Mmax prior distributions were developed from a study of SCR earthquakes. The process involved subdividing the SCR into tectonic domains on the basis of characteristics such as crustal extension, crustal age, age of major deformation, state of stress, and orientation of predominant structural grain relative to stress. A catalog of SCR earthquakes was developed in which moment magnitudes were assessed for each earthquake. The SCR earthquake catalog was then used to assess the largest observed earthquake in each domain. The mean and standard deviation of these values for specified subsets of the SCR domains provided normal distributions for $m_{\max-obs}$. The last step in converting the distribution for $m_{\max-obs}$ into a prior distribution for m^u was to make a bias correction to account for the fact that $m_{\max-obs}$ is in nearly all cases less than m^u. The bias correction is based on the assumption that the size distribution of earthquakes in a source region corresponds to a truncated exponential distribution between a specified minimum magnitude m_0 and the maximum magnitude for the region, m^u. Using this assumption, the cumulative probability distribution for the largest earthquake in a sample of size N given a specified value of m^u is (Johnston et al., 1996):

$$F[m_{\max-obs}] = \left[\frac{1-\exp\{-b\ln(10)(m_{\max-obs}-m_0)\}}{1-\exp\{-b\ln(10)(m^u-m_0)\}}\right]^N \quad \text{for } m_0 \le m_{\max-obs} \le m^u \quad (5.2.1\text{-}2)$$

Figure 5.2.1-3 shows median (50th percentile) values of $m_{\max-obs}$ as a function of m^u and N. These values were used to make a bias adjustment from $m_{\max-obs}$ to m^u by selecting a value of m^u such that the median value of the maximum observed, $\hat{m}_{\max-obs}$, equals the actual observed value. An example is illustrated on Figure 5.2.1-3 for the case in which the value of $m_{\max-obs}$ equal to 5.7 for a sample size of 10 earthquakes equal to or larger than m_0 of 4.5. The $N = 10$ curve shown on Figure 5.2.1-3 indicates that a m^u value of 6.3 produces a value of $\hat{m}_{\max-obs}$ equal to 5.7. Thus a magnitude of 6.3 represents the bias-adjusted value of m^u for the source.

The relationships shown on Figure 5.2.1-3 indicate that for a given value of N, the value of $\hat{m}_{\max-obs}$ reaches an asymptotic level as m^u increases. Thus it is not possible to make meaningful bias corrections for large values of $m_{\max-obs}$ unless the sample size is large. This fact led Johnston et al. (1994) to use domain pooling to increase the sample size. Tectonic domains that had the same characteristics (e.g., same type of crust, same age, same state of stress) were combined into a "superdomain" under the assumption that the common set of characteristics would correlate with a common value of m^u. The mean value of $m_{\max-obs}$ in a set of superdomains was then bias-adjusted to produce a mean value of m^u using the average sample size in the superdomains. This bias adjusted value of m^u along with the standard deviation of $m_{\max-obs}$ was used as a prior distribution for Mmax.

Given the importance of the prior distribution in most applications of the Bayesian Mmax approach, the CEUS SSC Project focused its efforts on development of updated prior distributions. This involved updating the SCR data sets and performing new statistical analyses.

The SCR data sets used to develop the Johnston et al. (1994) Mmax prior distributions consist of the SCR earthquake catalog and the set of SCR domains. The SCR earthquake catalog was updated for this study. The principal source for this update was the catalog published by Schulte and Mooney (2005) that extended the SCR catalog of Johnston et al. (1994) from the end of 1990 to the end of 2003 as well as added additional earthquakes for the earlier time period identified in the literature. Earthquakes occurring in SCRs for the period 2004 through 2008 were added from the Harvard Moment Tensor (HMT) catalog. All earthquakes were assigned moment magnitudes and magnitude uncertainties based on the values given in Johnston et al. (1994) and Schulte and Mooney (2003). These values were updated using the relationships and assessments presented in Johnston (1996a, 1996b) as well as data from the CEUS SSC catalog for earthquakes in the CEUS. Moment magnitudes from the HMT catalog were assigned a nominal uncertainty of 0.1-magnitude unit standard deviation. Moment magnitudes for specific earthquakes, such as the 1811-1812 New Madrid sequence and the 1886 Charleston earthquake, were assigned according to the current literature. Appendix K presents the updated SCR earthquake catalog.

Updating the basis for Johnston et al.'s (1994) division of the SCR crust into 255 tectonic domains was beyond the scope of this project. However, the assignments of age, stress state, and stress structure angles to the 255 SCR domains were reviewed and updated using recent information. Appendix Table K-2 lists the 255 SCR domains defined by Johnston et al. (1994) and the updated domain characteristics defined for each. The boundaries of these domains are included in the Project database.

The updated SCR catalog was then used to assess the value of $m_{\max-obs}$ for each domain and the sample size N representing the number of earthquakes **≥ M** 4.5. Following the approach of Johnston et al. (1994), an equivalent sample size corrected for completeness is obtained for each domain. Johnston et al. (1994) estimated catalog completeness intervals for various portions of the SCR. These were used to define the completeness-corrected equivalent earthquake count for each domain based on the catalog completeness period for magnitudes equal to $m_{\max-obs}$. The typical case is that for magnitudes smaller than $m_{\max-obs}$ the catalog completeness periods are shorter than those for $m_{\max-obs}$. Assuming that earthquake occurrence rates are stationary in time for the period covered by the catalog, then the maximum likelihood estimate of the rate of earthquakes of magnitude m_i is given by the number of earthquakes, $N(m_i)$ in the completeness period for m_i divided by the length of the completeness period, $T_C(m_i)$. An estimate of the total number of earthquakes of the smaller magnitude that would have been recorded in the completeness period for $m_{\max-obs}$ is then given by the estimated rate of occurrence for magnitude m_i multiplied by $T_C(m_{\max-obs})$:

$$N_{\text{Completeness Corrected}}(m_i) = N_{\text{In period } T_C(m_i)} \times \frac{T_C(m_{\max-obs})}{T_C(m_i)} \qquad (5.2.1\text{-}3)$$

The values of $N_{\text{Completeness Corrected}}(m_i)$ are then summed to provide the completeness-corrected sample size for each domain. These values are listed in Appendix Table K-2.

The updated domain data set was then used to assess appropriate prior distributions. The first step reviewed was the development of the superdomains by pooling domains with the same characteristics. Johnston et al. (1994) found that separation of the domains into extended and non-extended crust was the most meaningful initial classification. They then used other domain characteristics to develop superdomains within these two classes. For the non-extended crust domains, the characteristics used were crustal age, stress state, and whether the predominant structures within a domain are favorably oriented with respect to stress (labeled as source stress angle). These three attributes were used to group the 146 non-extended domains into a small number of superdomains. The characterization of the extended crust domains included additional factors such as the type of crust adjacent to the domain, the presence or absence of cross structures, and the presence or absence of multiple features. These additional characteristics were used by Johnston et al. (1994) to assign the 109 extended domains to superdomains. The result of using more characteristics to define extended crust superdomains was to increase the number of possible superdomains at the expense of reducing the sample size in each. Therefore, it was decided to use the same three characteristics—age, stress, and source stress angle—as the basis for defining both extended and non-extended superdomains. This provided a consistent basis for superdomain definition and also increased the average sample size in the extended superdomains. The result was 15 active (i.e., containing earthquakes) non-extended superdomains and 15 active extended superdomains. Figure 5.2.1-4 shows histograms of $m_{\max-obs}$ for the extended and non-extended superdomains.

The significance of the extended and non-extended superdomain classifications was examined by comparing the statistics of $m_{\max-obs}$ for the two classes. The results indicated that the extended and non-extended superdomain sites had similar mean values of $m_{\max-obs}$. Applying a Student's *t*-test (with the Welsh modification for unequal variances) to compare the two samples yielded a very high probability (*p*-value) of 0.99 that the two populations from which the samples were drawn (extended superdomains and non-extended superdomains) have the same mean. This suggested that the classification scheme for Mmax priors for extended and non-extended crust needed to be revisited.

Examination of the values of $m_{\max-obs}$ in the extended crust superdomain set suggested that the larger values of $m_{\max-obs}$ occurred in the domains with Mesozoic and younger ages. Accordingly, an alternative grouping of superdomains was tested in which extended superdomains of Mesozoic and younger ages formed one group, designated MESE, and the older extended and the non-extended superdomains formed the second group, designated NMESE. Figure 5.2.1-5 shows histograms of $m_{\max-obs}$ for these two sets of superdomains. Application of the Student's *t*-test to these two samples yielded a lower *p*-value of 0.46, suggesting that it is more likely that the two populations have different mean values.

Building on the concept that a relatively young age of extension is an important factor, a new age classification was assigned to each extended domain based on the age of the most recent

extension episode. These age assignments are listed in Appendix Table K-2. The resulting sets of superdomains containing the Mesozoic and younger extended superdomains (MESE) are listed in Table 5.2.1-1, and those containing older extended and non-extended superdomains (NMESE) are listed in Table 5.2.1-2. Figure 5.2.1-6 shows histograms of $m_{\max-obs}$ for the revised superdomains. Using these two classifications resulted in a further reduction in the *p*-value obtained from the Student's *t*-test to 0.31.

An additional step taken was to remove from the samples those superdomains for which the stress classification is "Unknown" as it is likely that they would be classified as either compressive or extensive with more data. The "Unknown" source–stress angle classification was retained because it includes cases where there is no clear structural grain in the domain. In addition, superdomains made up of a single domain or those with a single earthquake were removed, assuming that the information was too limited to assess $m_{\max-obs}$ for these cases. The set of superdomains used in the final test are flagged in Tables 5.2.1-1 and 5.2.1-2. Figure 5.2.1-7 shows histograms of $m_{\max-obs}$ for this final set of superdomains. Application of the Student's *t*-test to these two samples yielded a *p*-value of 0.09, indicating a reasonable likelihood that the mean $m_{\max-obs}$ values for the two populations are different.

The fact that the computed *p*-value for the two samples listed in Table 5.2.1-1 is not as low as conventional values used to indicate statistical significance (e.g., a value of 0.05) indicates that in fact the distinction between Mesozoic and younger extension and older extension and non-extension is not important to the assessment of Mmax. Therefore, a second grouping of domains into superdomains was performed using only age, stress, and source stress angle. The resulting composite superdomains (COMP) are listed in Table 5.2.1-3 and the assignment of each of the domains to these combined superdomains is indicated on Appendix Table K-2. Figure 5.2.1-8 shows a histogram of $m_{\max-obs}$ for the combined superdomains.

The remaining step in computing the priors was to compute the mean and standard deviation of $m_{\max-obs}$ and perform a bias correction to adjust the mean of $m_{\max-obs}$ to the mean of m^u. The SCR catalog contains many earthquakes whose moment magnitudes were estimated from macroseismic data and are, therefore, subject to considerable uncertainty. It is important to capture the impact of this uncertainty on the estimates of the mean and standard deviation of $m_{\max-obs}$. The process used to compute the statistics of $m_{\max-obs}$ for each group of superdomains is as follows.

The updated SCR catalog was used to construct an earthquake catalog for each superdomain. As discussed in Section 3.3.1, it is important to take into account the background distribution for earthquake size in computing the statistics of earthquake sizes. Assuming that earthquake size conforms to a truncated exponential distribution, then the density function for earthquake magnitude is exponential in shape— $f_M(m) \propto \exp(-\beta m)$, where β is the Gutenberg-Richter *b*-value expressed in natural log units [$\beta = b \times \ln(10)$]. If it is assumed that uncertainty in the observed value of magnitude can be represented by a normal distribution, $f_{\hat{M}|M} \sim N(M, \sigma)$, then using Bayes's Theorem, the distribution for the true size of an observed earthquake is proportional to the product of the normal and exponential distributions. The result is a normal

distribution centered on the expected magnitude $E[\mathbf{M}]$ given by Equation 3.3-5 (repeated below for ease of reference).

$$E[\mathbf{M}] = \hat{\mathbf{M}} - \beta\sigma^2[\mathbf{M}|\hat{\mathbf{M}}] \qquad (5.2.1\text{-}4)$$

In Equation 5.2.1-4, $\hat{\mathbf{M}}$ is the observed value of magnitude for the earthquake and σ is the standard deviation of the estimate. Thus, for the case of direct observation of moment magnitude from data, the effect is to shift the expected value of the true magnitude from $\hat{\mathbf{M}}$ to $E[\mathbf{M}]$ by the amount $-\beta\sigma^2$. However, as discussed in Section 3.3.1, the estimation of moment magnitude using regressions against other size measures (other magnitude scales or shaking intensity measures) directly produces the estimate $E[\mathbf{M}]$. This situation applies to the majority of the earthquakes in the SCR database. The magnitudes listed in Table K-1 are values of $E[\mathbf{M}]$. In the cases where the magnitudes were estimated from correlations with other size measures, then the values listed in Table K-1 are taken directly from the source. In the case where actual moment magnitudes are observed, then $E[\mathbf{M}]$ is obtained using Equation 5.2.1-4. The value of β used was estimated from the combined catalog of earthquakes in the superdomains, allowing for differences in the rate of activity and average catalog completeness among the superdomains.

Using the values of $E[\mathbf{M}]$ and $\sigma[\mathbf{M}]$ for earthquakes in the SCR database, a catalog of earthquakes associated with each superdomain was simulated and the statistics of $m_{\max-obs}$ for the superdomain set was computed. The simulation process was repeated 10,000 times to generate a composite estimate of the mean and standard deviation of $m_{\max-obs}$ for the three sets of superdomains listed in Tables 5.2.1-1, 5.2.1-2, and 5.2.1-3. The results are listed in Table 5.2.1-4. The mean values of $m_{\max-obs}$ were then adjusted to a mean value of m^u by applying the bias correction process given by Equation 5.2.1-2 and illustrated on Figure 5.2.1-9. The average sample sizes and *b*-values for each group of superdomains are also listed in Table 5.2.1-3. Figure 5.2.1-9 shows the three bias adjustment curves and Table 5.2.1-4 lists the resulting values of mean m^u. After evaluation the results of the above analysis, as well as considering the meaning of prior distributions, it was assessed that the uncertainty in assigning prior distributions for maximum magnitude to the CEUS sources should incorporate two alternative conceptual models: a "two-priors" distribution model and a "single-prior" (composite) distribution model. These alternatives are described below.

5.2.1.1.1 Two-Priors Distribution Model

The results of the updated statistical analysis show that the most significant separation of the SCR database is between superdomains that show Mesozoic-and-younger extension and those that do not. The data used are based on the domains in the SCR database, defined by their most recent age of extension and known characteristics of tectonic stress, and on those superdomains with more than one domain and more than a single earthquake. This subdivision shows reasonable statistical significance (i.e., there is a relatively low *p*-value of 0.09, indicating that based on the difference in the mean values, there is a low probability that the two data sets came

from the same data population), but the statistical significance is not strong. Using the bias-adjusted values listed in Table 5.2.1-4, the Mmax priors for the "two-prior" model are as follows:

	Mean Mmax	Sigma
Mesozoic and younger extension (MESE):	7.35	0.75
Non-Mesozoic and younger extension (NMESE):	6.70	0.61

These prior distributions are assigned to the applicable Mmax zones of the CEUS SSC study area, whose boundary and uncertainty have been appropriately accounted for in the SSC model (Section 6.2.1). Also, each seismotectonic zone is assigned one of these prior distributions (or both are assigned with associated weights) to obtain its appropriate prior distributions.

5.2.1.1.2 Composite-Prior Distribution Model

This model assumes that a single, composite-prior distribution is appropriate throughout the region covered by the CEUS SSC seismic source model. The composite prior represents the concept that the characteristics within SCRs that control Mmax cannot be identified with confidence, despite our considerable effort to do so. Therefore, the only distinction made is that the applicable global database must be from SCRs (as defined in Johnston et al., 1994), but no subdivisions are imposed beyond that. This alternative conceptual model is consistent with the relatively weak statistical significance of any subdivision of the SCR as indicators of Mmax. Using the values listed in Table 5.2.1-4, the prior distribution for the composite data set is as follows:

	Mean Mmax	Sigma
Composite (COMP):	7.20	0.64

5.2.1.1.3 Relative Weights

The "two-priors" model is assessed to have slightly higher weight (0.6) than the "composite-prior" model (0.4). Higher weight reflects the fact that there is some statistical significance to the separation of the data into the two Mmax zones, as well as an expectation that the locations of geologically recent crustal extension should have ample faults of sufficient dimensions for larger earthquakes. However, a stronger weight than 0.6 is not assessed because the statistical significance of the separation is not strong and could only be evaluated after multiple domains were combined into superdomains to increase sample size for statistical evaluation.

5.2.1.1.4 Truncation of the Prior Distributions

The Mmax prior distributions described above are represented by unbounded normal distributions. However, there is likely an upper limit to the size of earthquakes that can occur in the CEUS and a bound on the minimum size of the maximum magnitude, as well.

The upper tail of the maximum magnitude distributions for all distributed seismicity seismic sources is truncated at **M** 8¼. This value is selected primarily on the basis of empirical data. **M** 8

¼ envelopes the largest magnitude events ever recorded in SCRs. The largest events reported in the SCR earthquake database include the 1668 **M** 7.87 Yellow Sea China earthquake, the **M** 7.8 New Madrid earthquake, and the 1819 M 7.8 Kutch earthquake (Appendix K). **M** 8¼ also approximates the largest crustal earthquakes recorded anywhere in the world outside of subduction zones, including the 1905 Mw 8.1-8.5 Bulnay earthquake (375 km rupture length) (Prentice et al., 2010), the 1949 **M** 8.1 Queen Charlotte earthquake (300–490 km rupture length) (Bostwick, 1984; Rogers, 1986), and the 1957 **M** 8.0–8.1 Gobi Altay earthquake (260 km rupture length; Prentice et al., 2010). The estimated rupture lengths for these earthquakes are comparable to or significantly longer than the rupture lengths of the largest strike-slip crustal earthquakes included in the Wells and Coppersmith (1994) and Hanks and Bakun (2002) worldwide databases that are used to develop empirical relationships for magnitude and fault rupture parameters. The Bulnay earthquake is poorly recorded, and estimates of magnitude are based on physical dimensions of the rupture (**M** 8.1, Schwartz et al., 1998) and analysis of historical Wiechert seismograms (**M** 8.2–8.5; Schlupp and Cisternas, 2007). Published magnitude estimates for the 1957 Gobi Altay earthquake range from M 7.8–8.3 (Prentice et al., 2010), with a preferred value of M 8.1 (USGS Earthquake Database, 2010; Choi et al., 2009). Based on these empirical observations of the largest crustal events that have occurred anywhere in the world, our estimate of **M** 8¼ as the upper-bound truncation of the maximum magnitude distribution reasonably captures the epistemic uncertainty in maximum magnitude. This value is larger than the magnitude of any observed SCR earthquakes.

The prior distributions are also truncated at the minimum magnitude that is assessed to be technically defensible, given a consideration of the observed magnitudes of earthquakes that have occurred with SCRs worldwide. Review of the values of $m_{\max-obs}$ listed in Tables 5.2.1-1, 5.2.1-2, and 5.2.1-3 for the superdomains used to develop the prior distributions indicates a minimum value of 5.5. Conceptually, this value should be increased slightly by the bias adjustment described above. However, bias adjustments are typically small at these low magnitudes (Figure 5.2.1-9). In addition, the parameter being assessed is the minimum of a distribution, which should be below the minimum observed. Therefore, it was assessed that an appropriate lower truncation point for the prior distributions is **M** 5½.

5.2.1.1.5 Addressing Criticism of the Bayesian Approach

Kijko et al. (2009) have criticized the use of the Bayesian approach for estimation of Mmax as producing results that are biased low. They demonstrate the issue by performing sumilations of catalogs of earthquakes generated from an exponential distribution with a known Mmax (set at the mean of a hypothetical Mmax prior distribution) and then estimating the Mmax using the posterior distribution generated from the product of the prior with the sample likelihood function (e.g., the lower right-hand plots on Figures 5.2-1 and 5.2-2). Discussions with the lead author (A. Kijko, pers. comm., 2011) indicate that Kijko et al. (2009) use the mode of the posterior as the point estimate of Mmax. However, the application of the Bayesian Approach for Mmax estimation in the CEUS SSC Project uses the full posterior distribution. A better point estimate of this result is the mean of the posterior rather than the mode. The effect of the use of the mode versus the mean can be clearly seen by comparing the posterior distributions on Figures 5.2.1-1 and 5.2.1-2. On Figure 5.2.1-1, the sample size is small and the posterior distribution has a shape very similar to the prior with only a minor shift in the mode. The shift in the mean between the

prior and the posterior is even less. On Figure 5.2.1-2, the sample size is large and the likelihood function has dramatically affected the shape of the posterior such that the mode occurs at the maximum observed value. However, the posterior distribution does extend to magnitudes larger than the maximum observed such that the mean would be a larger value.

The effect of use of the mode of the posterior, as in Kijko et al. (2009), versus the mean of the posterior, a more appropriate point estimate of the results used in the CEUS SSC Project, is illustrated on Figure 5.2.1-10. A prior distribution for maximum magnitude with a mean of 6.9 and a standard deviation of 0.5 was assumed. Various size samples of exponentially distributed earthquakes were then simulated from an exponential distribution of magnitudes truncated at the mean of the prior, 6.9. For each sample the likelihood function was used to develop a posterior distribution. The mean and the mode of this posterior were then computed. The process was repeated for 1000 simulations of each sample size. The average values of the mean and mode for each sample size are plotted on Figure 5.2.1-10. These results confirm the conclusion of Kijko et al. (2009) that the mode of the posterior is a biased estimate of Mmax, with the bias to smaller values. However, the mean of the posterior does not display any significant bias. Therefore, it is concluded that the application of the Bayesian approach using the full posterior distribution should not lead to biased estimates of Mmax.

5.2.1.1.6 Application to CEUS Distributed Seismicity Sources

The application of the Bayesian Mmax approach requires development of the likelihood function Equation 5.2.1-1 from the assessment of the maximum observed earthquake in each source and the number of earthquakes larger than a specified minimum magnitude. The assessment of these parameters was based on the project earthquake catalog (Chapter 3), and descriptions of the largest earthquakes in each source are provided in Chapters 6 and 7. The minimum magnitude used for the earthquake counts is **M** 4.7, as this is at the lower limit of one of the magnitude intervals used in the earthquake catalog analysis and there is a focus on earthquakes of significance for a hazard analysis. The magnitude estimates for the earthquakes in the CEUS catalog are subject to uncertainty. This uncertainty was incorporated into the assessment of $m_{\max-obs}$ using the procedure described above for the SCR domains by simulating 100,000 sets of earthquake magnitudes for the domain catalogs, selecting the largest magnitude for each simulated catalog, and then binning these values in 0.1-magnitude units to develop a distribution for $m_{\max-obs}$. In some cases, there is uncertainty in assigning the largest observed earthquake to a particular source, due to uncertainty in location for earthquakes near source boundaries. This uncertainty was incorporated by developing alternative catalogs including or excluding the particular earthquake, developing simulated distributions for $m_{\max-obs}$ for each case, and then combining the results to produce a composite distribution.

The uncertainty distribution for $m_{\max-obs}$ was incorporated into the assessment of Mmax distributions for the distributed seismicity sources by developing a likelihood function for each possible value of $m_{\max-obs}$ using Equation 5.2.1-1, using this likelihood function with the appropriate prior to develop a posterior Mmax distribution, and then weighting this posterior by

the probability assigned to the value of $m_{\max-obs}$. The sum of these weighted posteriors represents the composite Mmax distribution for the source.

5.2.1.2 Kijko Approach to Mmax Assessment

The Kijko approach (Kijko, 2004) provides an assessment of Mmax based on only the observed distribution of earthquakes in a region. As described in Kijko (2004), the cumulative distribution for $m_{\max-obs}$ in a source zone is given by

$$F_{m_{\max-obs}}(m) = \begin{cases} 0 & m < m_0 \\ \left[F_M(m)\right]^N & m_0 \le m \le m^u \\ 1 & m > m^u \end{cases} \quad (5.2.1\text{-}5)$$

where $F_M(m)$ is the cumulative probability distribution for magnitude and parameters m_0, m^u, and N as defined above in Section 5.2.1.1. Kijko (2004) shows that the expected value of $m_{\max-obs}$ is given by the expression

$$E(M_{\max-obs}) = \int_{m_0}^{m^u} m \frac{dF_{\max-obs}(m)}{\mathrm{d}m} \mathrm{d}m \quad (5.2.1\text{-}6)$$

Integration by parts yields

$$E(M_{\max-obs}) = m^u - \int_{m_0}^{m^u} F_{m_{\max-obs}}(m) \mathrm{d}m \quad (5.2.1\text{-}7)$$

which leads to the expression:

$$m^u = E(M_{\max-obs}) + \int_{m_0}^{m^u} F_{m_{\max-obs}}(m) \mathrm{d}m \quad (5.2.1\text{-}8)$$

Kijko (2004) presents three alternative relationships for the application of Equation 5.2.1-8. The first is based on application of the truncated exponential distribution for $F_M(m)$, leading to the distribution for $F_{m_{\max-obs}}(m)$ given by Equation 5.2.1-2. This is designated the K-S (Kijko-Sellevoll) estimator by Kijko (2004). The second relationship incorporates uncertainty in β (the *b*-value) and takes the form

$$F_{m_{\max-obs}}(m) = \begin{cases} 0 & m < m_0 \\ \left[C_\beta \left\{ 1 - \left(\dfrac{p}{p + m - m_0} \right)^q \right\} \right]^N & m_0 \le m \le m^u \\ 1 & m > m^u \end{cases} \quad (5.2.1\text{-}9)$$

where $p = \overline{\beta} / \sigma_\beta^2$, $q = \left(\overline{\beta} / \sigma_\beta\right)^2$, and $C_\beta = \left[1 - \left\{p / (p + m^u - m_0)\right\}^q\right]^{-1}$. This second formulation is designated the K-S-B (Kijko-Sellevoll-Bayes) estimator by Kijko (2004). The third formulation makes no assumption about the form of the magnitude distribution. Instead, it used a kernel density estimation technique to estimate the form of the magnitude distribution from the observed magnitudes. This form, designated the N-P-G form in Kijko (2004) was not investigated for application in the CEUS SSC. Its intended use is to cases where characteristic earthquake type behavior is observed. In the CEUS SSC model, known earthquakes of this type are modeled by a separate set of RLME sources. With these earthquakes addressed by separate sources, the remaining earthquakes in the Project catalog display an approximately exponential distribution, which are addressed by the K-S and K-S-B estimators.

The behavior of the K-S and K-S-B estimators as a function of $m_{\max-obs}$ and *N* is shown on Figure 5.2.1-11. The values of m^u were computed by iteratively solving Equation 5.2.1-8 until convergence. Note that the estimates for the K-S estimator reach a point where the iterative solutions do not converge. The K-S-B estimator converges for all values of m^u tried. Also shown on Figure 5.2.1-11 is the relationship between m^u and the median value given by Equation 5.2.1-2, indicating that the bias adjustment used in developing the Bayesian priors is nearly equivalent to the Kijko (2004) estimators of m^u. This is not surprising given that they are based on similar formulations.

The relationships shown on Figure 5.2.1-11 indicate that the Kijko estimators for the expected value of m^u are subject to the same need for large values of *N* as the bias adjustment relationship used to develop the Bayesian priors. In order to apply the approach to situations where *N* is small, an additional constraint must be used. This constraint is supplied by the assumption of a limiting size for CEUS earthquakes that was used to truncate the Bayesian prior distributions in Section 5.2.1.1. Kijko (2004) presents the following relationship for the cumulative probability function for m^u:

$$P(m^u < z) = F(m^u) = 1 - F_{m_{\max-obs}}(m_{\max-obs} | z) \quad (5.2.1\text{-}10)$$

Figure 5.2.1-12 shows the behavior of Equation 5.2.1-10 as a function of *N*. For values of *N* below about 100, the value of $F(m^u)$ reaches a limiting value less than 1.0 as m^u approaches infinity (∞). Kijko (2004) discusses this limiting value of $F(m^u)$ as an indicator of the probability that the K-S or K-S-B estimator provides meaningful results.

As indicated by Figure 5.2.1-12, the cumulative distribution function $F(m^u)$ does not provide a complete probability distribution for m^u in the unbounded case. However, if a limiting value for m^u is imposed, then $F(m^u)$ can be truncated at this point and renormalized to produce a probability distribution for m^u.

While this approach allows one to construct a proper probability distribution for m^u regardless of the value of N and $m_{\max-obs}$, it involves truncation of substantial probability mass for most practical applications. The truncated probability mass is given by $P(m^u > 8¼)$ calculated using Equation 5.2.1-10. The assessment was made that the point where the truncated probability mass equals the remaining probability mass, $P(m^u > 8¼) = 0.5$, represents the limit of applicability of the Kijko Mmax approach.

The application of a lower truncation point for Mmax distributions at **M** 5½ requires a modification of the approach. The first step is to compute the unbounded cumulative distribution function $F(m^u)$ using Equation 5.2.1-10. This distribution is modified by imposing the requirement that m^u must be ≥5½ by computing its derivative, removing the probability mass below 5½, if any, and then renomalizing. The effective result is to increase the value of $P(m^u > 8¼)$ for those cases where the maximum observed magnitude is less than 5½.

After a variety of exploratory calculations and considerations, the following key constraints were made in the application of the Kijko approach to assessment of Mmax for the CEUS SSC model:

- The K-S-B approach (Kijko, 2004) that incorporates the uncertainty in *b*-value for the exponential distribution will be used, with an upper truncation at M 8¼ and a lower truncation at M 5½.

- Uncertainties in the magnitude of the largest earthquake in each source are assessed using the approach described in Section 5.2.1.1.5.

- For those sources with a maximum observed magnitude derived from the paleoseismic record, the Kijko approach cannot be used because the approach assumes a complete record (historical earthquakes are adjusted for catalog completeness). We have no approach to formally evaluate the completeness of the paleoseismic record. An exploratory calculation using, for example, the four earthquakes identified for the Illinois Basin Extended Basement (IBEB) source, which has the largest number of observed paleoseismic earthquakes for any seismic source in the model, confirms that a high value of $F(m^u > 8¼)$ results from consideration of these events. Therefore, giving zero weight to the Kijko approach is reasonable.

- The weighting scheme for the use of the Bayesian approach versus the Kijko approach should be source-specific (to account for the numbers and magnitudes of observed earthquakes). In no case should the Kijko approach be given higher weight than the Bayesian approach, considering the higher reliance on the Bayesian approach (and similar analogy-based approaches) by the technical community.

5.2.1.3 Weights for the Alternative Mmax Approaches

Because the probability that Mmax is much greater than 8 is a measure of the statistical stability of the Kijko approach—and is noted by its author as providing a measure of whether the approach is producing meaningful Mmax estimates—it is used as an index for weighting the Kijko versus Bayesian approaches. At values of $P(m^u > 8¼) = 0.5$ and higher, the Kijko approach is assessed to have zero weight. Likewise, at $P(m^u > 8¼) = 0$, where Kijko's approach is at its strongest with large numbers of larger earthquakes, the Kijko approach is assessed to have equal weight (0.5) relative to the Bayesian approach. With these endpoints, a linear relationship is developed that provides the weight for the Kijko approach as a function of the calculated value of $P(m^u > 8¼)$ for the source of interest.

Using the linear relationship between $P(m^u > 8¼)$ values and the Kijko weights, the calculated weights for the Kijko approach for the various source zones are shown in Table 5.2.1-5. Using the assumptions discussed previously and the cutoff of a $P(m^u > 8¼)$ value equal to or greater than 0.5, it can be seen that the Kijko approach is given zero weight for several sources: for some sources, the Kijko approach is given zero weight because the paleoseismic record is used to constrain the size of the maximum observed event; for others, zero weight is given because of the high $P(m^u > 8¼)$ values.

5.2.1.4 Example Mmax Distributions

Using the Bayesian and Kijko approaches in the manner described in this section, Mmax distributions were developed for both the Mmax zones (Section 6.2.3) and for the seismotectonic zones (Section 7.4.2). An example of how these distributions are displayed is given on Figures 5.2.1-13 and 5.2.1-14. As a reminder, in the master logic tree, the Mmax zones approach is assessed a weight of 0.4, versus 0.6 for the seismotectonic zones approach. Given the Mmax zones approach, the weight assessed for a single, composite zone is 0.4, versus 0.6 for two Mmax zones. Given two Mmax zones, the weight assessed for the "narrow" interpretation of Mesozoic and younger extended zone (MESE-N) configuration is 0.8, versus a weight of 0.2 assessed for the "wide" interpretation (MESE-W).

Figure 5.2.1-13 shows the Mmax distribution for the Mesozoic and younger extended zone (MESE-N), which is one of the Mmax zones. Shown in all of the plots, regardless of the assessed weights, are the results for the Kijko approach, the Bayesian approach using the composite prior distribution, and the Bayesian approach using a MESE or non-MESE prior distribution. The weights assessed for the Kijko approach in Table 5.2.1-5 are functions of the $P(m^u > 8¼)$ value and whether paleoseismic data is used to define the size of the maximum observed magnitude. As discussed in Section 5.2.1.1.3, the weight assessed for the "two-priors" model is 0.6, and to the "composite" model, 0.4. Mmax estimates for each of these prior distributions are given along with their weights, which are the product of the Kijko versus Bayesian weights and the two-priors versus composite-prior weights. The final composite Mmax distribution is shown by the thick red histogram that reflects the probability mass function at 0.1-magnitude increments.

A second example of Mmax distribution is shown on Figure 5.2.1-14 for the Northern Appalachian (NAP) seismotectonic zone. The same convention is used in the labeling and

display of the results for the Kijko and Bayesian approaches. In this case, the weight assessed for the Kijko approach is zero, based on a $P(m^u > 8¼)$ being greater than 0.5. The results of the Kijko approach are shown simply for purposes of comparison, but they are not incorporated into the final composite Mmax distribution, shown in red.

For use in the hazard calculations, the maximum magnitude distributions for each source are then represented by a five-point discrete approximation to an arbitrary continuous distribution developed by Miller and Rice (1983). (The use of a discrete representation of a continuous distribution is described in more detail in Section 5.3.3.1.3.) The distributions for the two example sources are listed in Table 5.2.1-6.

5.2.2 Other Mmax Issues

One of the key uncertainties regarding the assessment of Mmax for seismic sources in the CEUS is spatial variation of Mmax throughout any given seismic source. That is, is it realistic to assume that a single Mmax distribution developed for a seismic source applies to all locations within that source? To a large extent, this issue is resolved within more active regions where active faults have been identified and Mmax can be assessed on a fault-specific basis using approaches such as fault rupture dimensions. In SCRs like the CEUS, however, the physical bases for specifying differences in Mmax spatially are not known with confidence. Unless specific faults can be identified that are the causative structures, the use of fault dimensions to constrain Mmax is problematic. The tectonic history of a region and the presence of significant geologic structures suggest that such structures might host future larger earthquakes, but no systematic studies have been conducted to show that the presence or absence of older geologic structures leads to differences in Mmax potential. Further, many SCRs show evidence of significant older geologic structures, but they show no evidence of large earthquake occurrence (Johnston et al., 1994). Also, in his inventory and evaluation of possible Mmax approaches, Wheeler (2009, pp. 12-13) cites the difficulties in using local geologic structures to estimate Mmax when there is no direct geologic evidence for young faulting.

Using current approaches, the broad distributions of Mmax developed for seismic sources in the CEUS are assumed to be entirely epistemic and applicable throughout the source of interest. If the spatial variation of Mmax were known, that is, if it were possible to say that one part of the source should have a higher Mmax than another part, then that variation could be captured as aleatory variability. The result might be two epistemic distributions—one for regions marked by higher Mmax and one for regions with lower Mmax—and both distributions could be narrower than the current distribution for the entire source. At present, if there is a basis for identifying a spatial variation in Mmax, that information is used in the CEUS SSC Project to identify a separate seismic source (see Section 4.2.3). The Mmax distributions for the seismic sources in the CEUS developed for the Mmax zones (Section 6.3.2) and for the seismotectonic zones (Section 7.4.2) are quite broad in many cases to reflect the current uncertainties that exist with the estimation of Mmax within an SCR. It is hoped that the estimation of Mmax will continue to be a focus of research efforts in the scientific community, and that those efforts will lead to reductions in the epistemic uncertainties that currently exist.

5.3 Earthquake Recurrence Assessment

5.3.1 Smoothing to Represent Spatial Stationarity

The CEUS SSC model is based to a large extent on an assessment that spatial stationarity of seismicity will persist for time periods of interest for PSHA (approximately the next 50 years for engineered structures). Stationarity in this sense does not mean that future locations and magnitudes of earthquakes will occur exactly where they have occurred in the historical and instrumental record. Rather, the degree of spatial stationarity varies as a function of the type of data available to define the seismic source. RLME sources are based largely on paleoseismic evidence for repeated large-magnitude ($M \geq 6.5$) earthquakes that occur in approximately the same location over periods of a few thousand years. Uncertainties in the locations and sizes of these events are a function of the types of data available (e.g., fault-specific repeated displacements on the Meers fault, and shaking effects over a relatively small region for the Marianna source). Because the record that defines the RLME sources spans a relatively long time period and records large-magnitude events, repeated events for these sources are expected to occur within a restricted location defined by the RLME source.

On the other hand, patterns of seismicity away from the RLME sources within the Mmax and seismotectonic zones are defined from generally small- to moderate-magnitude earthquakes that have occurred during a relatively short (i.e., relative to the repeat times of large events) historical and instrumental record. Thus the locations of future events are not as tightly constrained by the locations of past events as for RLME sources. As discussed below in Section 5.3.2, the spatial smoothing operation is based on calculations of earthquake recurrence within one-quarter-degree or half-degree cells, with allowance for "communication" between the cells. Both *a*- and *b*-values are allowed to vary, but the degree of variation has been optimized such that *b*-values vary little across the study region. Also, *a*-values are neither "spiky," reflecting too strong of a reliance on the exact locations and rate densities of observed events, nor too smooth, reflecting the belief that the observed record does not provide a spatial constraint on rate density variation. Likewise, the recurrence calculation considers weighting of magnitudes in the recurrence rate calculations; thus moderate events are assigned more weight than smaller events.

Ultimately, any smoothing operation of seismicity is founded on expert judgment related to the expected future distribution of moderate- to large-magnitude events. The CEUS SSC model gives strong consideration to locations of RLME sources as being spatially stationary through the future time periods of interest. The smoothing operation within the distributed seismicity zones results in variations in *a*- and *b*-values over scales that were judged by the TI Team to be reasonable, given the technical community's views on spatial stationarity and the relationship between observed small- to moderate-magnitude seismicity and future moderate- to large-magnitude seismicity.

5.3.2 Smoothing Approach

5.3.2.1 Development of Penalized-Likelihood Approach and Formulation

5.3.2.1.1 Model for the Penalized-Likelihood Function of Recurrence Parameters

This section introduces and discusses the formulation for the calculation of earthquake recurrence parameters for one source zone (or one cell within a source zone) under very simple assumptions, then reformulates and expands the formulation to include a number of elements that make the formulation more robust, realistic, and flexible. These elements include the reformulation in terms of magnitude bins, and the introduction of magnitude-dependent weights, catalog incompleteness, the effect of maximum earthquake magnitude (Mmax), spatial variation of parameters within the source zone, and the prior distributions of *b*.

The first model to consider is the basic model, first presented by Aki (1965) and others. In this formulation, the number of earthquakes with magnitudes in the interval between *m* and *m* + *dm* that occur in a source zone, within a time period of duration *T*, follows a Poisson probability distribution, with mean given by the following expression:

$$n(m \le M < m + dm) = ATv\beta e^{-\beta(m-m_0)}dm \qquad m \ge m_0 \qquad (5.3.2\text{-}1)$$

where A is the area of the source zone (in units of equatorial degrees squared), T is the time period of interest (i.e., the duration of the catalog in years), v is the rate per unit time and per unit area for earthquakes with magnitude greater than or equal to m_0 (the lowest magnitude considered in the recurrence analysis), and β is the slope of the exponential magnitude-recurrence law (i.e., the *b*-value times ln[10]). In addition, it is assumed that the number of earthquakes in non-overlapping magnitude intervals is independent, so that earthquake occurrences in time represent a marked Poisson process.

Before going further, it is useful to make some comments about the above formulation. First, Mmax is not considered at this early stage. Second, the magnitude m_0 is not the same as the lower-bound magnitude considered in the hazard integration; it is typically much smaller. This project uses m_0 equal to moment magnitude 2.9. Finally, many of the assumptions in this initial formulation will be relaxed later in this section. Still, this initial formulation serves as a building block for the more general formulation and is helpful because it lends itself to the development of useful insights about the methodology.

Data input to this first simplified model consist of the list of earthquake magnitudes in the catalog; i.e., $m_1, m_2, \ldots m_N$, where N is the number of earthquakes in the catalog. In this project, these magnitudes are all moment magnitudes, and some of them may have been converted to moment magnitude from the magnitude, epicentral intensity, or felt area in the primary catalogs (see Section 3.3). In addition, the catalog is considered complete over the time period T, for all magnitudes above m_0. A somewhat surprising feature of the Poisson process is that the occurrence times of these earthquakes are immaterial, as long as the earthquakes occurred within

the time period *T* of the catalog. This property is actually a natural consequence of the "lack of memory" property of the Poisson process.

The likelihood function for the model parameters takes the form

$$\ell(\nu,\beta)=\prod_{i=1}^{N}(\nu\ AT\beta e^{-\beta(m_i-m_0)})\exp\left[-\nu A\int_{m_0}^{\infty}T\beta e^{-\beta(m-m_0)}dm\right] \qquad (5.3.2\text{-}2)$$

where all multiplicative constants that are independent of the parameters have been omitted because they do not affect the results. In this equation and in the equations that follow, index *i* denotes the earthquakes in the catalog and *N* denotes the number of such earthquakes. Note that the integral is equal to *T* in this special case because *T* is independent of magnitude (and can be moved out of the integral) and the exponential probability density function integrates to unity. This simplification does not apply in the more realistic case to be considered later, where the completeness time depends on magnitude.

In general, the likelihood function indicates the degree of consistency between the parameters that one wants to estimate (in this case, ν and β) and the available data (in this case, the number of earthquakes and their magnitudes during a time period *T*). If the data are abundant, the likelihood function has a narrow shape, indicating low uncertainty. If the data are scarce, the likelihood function has a broad shape, indicating high uncertainty. In the context of the maximum-likelihood method, one chooses the parameters that maximize the likelihood function and one estimates the uncertainty using the second derivatives of the natural log of the likelihood, evaluated at the point of maximum likelihood.

To derive Equation 5.3.2-2, one can formulate the Poisson likelihood function for each small interval of width dm, using Equation 5.3.2-1, multiply these likelihoods, and then compute the limit of this product as dm approaches 0, taking into account that when dm is very small, the number of earthquakes in an interval is either 0 or 1.

Note that the integral in Equation 5.3.2-2 is equal to T in this special case because T is independent of magnitude (and can be moved out of the integral) and what remains inside the integral (i.e., the the exponential probability density function) integrates to unity. This simplification does not apply in the more realistic case to be considered later, where the completeness time depends on magnitude. If we apply this simplification, we note that the likelihood function is separable into a product of factors that depend on the rate ν and factors that depend on the exponential slope β, i.e.,

$$\ell(\nu,\beta)=(\nu AT)^N e^{-\nu AT}\times\prod_{i=1}^{N}(\beta e^{-\beta(m_i-m_0)}) \qquad (5.3.2\text{-}3)$$

Note that the ν -dependent portion and the β-dependent portions of the likelihood function are separable, This separability is lost later, as the formulation is made more robust and realistic.

The first (ν -dependent) portion in the above equation is the traditional likelihood for a Poisson distribution. The second (β -dependent) portion is the traditional likelihood for an exponential distribution.

This separability also implies that the maximum-likelihood estimates of ν and β are decoupled in this simple case. In particular, by taking the logarithm of the above expression, differentiating with respect to each parameter, making the result equal to 0, and solving for the parameter, we obtain the following well-known results for the maximum-likelihood estimates of ν and β:

$$\nu_{ML} = \frac{N}{AT} \tag{5.3.2-4}$$

$$\frac{1}{\beta_{ML}} = \frac{1}{N}\sum_{i=1}^{N}(m_i - m_0) \tag{5.3.2-5}$$

The last result was first derived independently by Aki (1965) and Utsu (1965). Note that in the case where $N = 0$, the maximum-likelihood estimate of the rate v is zero. On the other hand, if one is working with the entire likelihood function (in order to represent an epistemic uncertainty), one sees that the v-dependent portion of the likelihood in Equation 5.3.2-3 for $N = 0$ follows an exponential distribution, which happens to have a mean value of $1/(AT)$ (i.e., the value in Equation 5.3.2-4 corresponding to $N = 1$). Thus, consideration of the full-likelihood function for ν gives rise to a natural "floor" in regions of low seismicity.

Note also that this likelihood function for the Poisson portion has sharper peaks (i.e., lower coefficient of variation, or COV) when N is large, indicating that only a narrow range of values of ν are consistent with the catalog. In contrast, the distribution is broad when N is small (becoming as broad as the exponential distribution when $N = 0$). This is illustrated on Figure 5.3.2-1.

Similarly, Figure 5.3.2-2 shows the likelihood function for the b-value, for different values of N. Again, the COV decreases as N increases.

Equation 5.3.2-3 is in fact easier to derive than the equivalent 5.3.2-2. One begins by introducing the standard assumptions that the number of earthquakes follows a Poisson distribution, independent of the magnitudes, and that the magnitudes are exponential and independent identically distributed (iid). Equation 5.3.2-3 then follows automatically as the product of the Poisson probability mass function and the N exponential probability density functions (again, omitting multiplicative constants that are independent of the parameters ν and β).

This simple formulation of the magnitude-recurrence model is useful as a starting point but is not always robust and sufficient in practice. The remainder of this section describes the various modifications to the formulation, in order to make it more robust and flexible.

The first difficulty arises from the continuous nature of the above formulation, which assumes that the exponential assumption applies at both small and large scales. In practice, the exponential assumption may not apply at small scales (i.e., over differences of a few tenths of a magnitude units, or less), for a number of reasons. Most importantly, some of the earthquake magnitudes in the catalog have been effectively binned into integer epicentral intensity and then converted to magnitude. Other possible causes include problems in the conversion to moment magnitude, edge effects, or the phenomenon investigated by Lombardi (2003).

This deviation from exponentiality is illustrated by Figure 5.3.2-3, which shows a histogram of the magnitudes in the earthquake catalog developed in Chapter 3 and used later in this section for the calculation of recurrence parameters. Focusing on the left portion of this figure (say, magnitudes lower than 3.5), one notices that the data deviate substantially from a smooth exponential shape. These deviations are not consistent with the exponential model, given the very large earthquake counts in these bins. These large earthquake counts mean that the continuous formulation of the likelihood may incorrectly interpret the local slope of the histogram around magnitude 3 as containing information about the exponential slope β.

In order to avoid these potential problems, the likelihood function is reformulated in terms of the earthquake counts in discrete magnitude bins, with bin sizes appropriately selected so that magnitudes converted from intensity fall in the middle of these bins.

Considering bins of δ magnitude units (i.e., first bin between magnitudes m_0 and $m_0 + \delta$, second bin between $m_0 + \delta$ and $m_0 + 2\delta$, etc.), one can write the likelihood function

$$\ell(\nu, \beta) = \prod_{k=1}^{\infty} (\nu \ ATp_k)^{n_k} \exp\left[-\nu A \sum_{k=1}^{\infty} Tp_k \right] \tag{5.3.2-6}$$

where n_k is the number of earthquakes in the *k*-th magnitude bin ($m_0^k \le m < m_0^{k+1}$),

$$p_k = p_k(\beta) = \exp\left[-\beta(m_0^k - m_0) \right] - \exp\left[-\beta(m_0^{k+1} - m_0) \right] \tag{5.3.2-7}$$

is the probability mass associated with the k-th magnitude bin, and $m_0^k = m_0 + (k-1)\delta$ is the magnitude at the lower end of the k-th magnitude bin. One can obtain Equation 5.3.2-6 from Equation 5.3.2-2 by replacing each occurrence of the exponential density $\beta \exp[-\beta(m - m_0)$ with the value of p_k for the magnitude bin where m falls. By working with earthquake counts in discrete bins, sensitivity to small-scale deviations from the exponential model (i.e., deviations at scales of of a few tenths of a magnitude unit or smaller) is removed, resulting in a more robust formulation. We note that the expression for the likelihood function in Equation 5.3.2.6 is is still a separable function of ν and β, although one cannot easily solve for β_{ML}.

Perhaps the best known example of this discrete-magnitude formulation for the likelihood function is the paper by Weichert (1980). The discrete-magnitude formulation was also used in the EPRI-SOG project.

It is also worth noting that the number n_k in Equation 5.3.2-6 is not necessarily an integer number. This project accounts for the effect of uncertainty in magnitude by calculating an "event factor" for each earthquake (see Chapter 3). Thus, n_k represents the sum of the event factors for all the earthquakes in the *k*-th magnitude bin.

A second modification is required because the assessment may result in a lower weight for lower magnitude bins. For instance, the magnitude-recurrence law may deviate from exponential, or the magnitude-conversion models or completeness model may be less reliable for lower magnitudes. The USGS (Petersen et al., 2008) use of alternative low-magnitude cutoffs may be viewed as a special case of magnitude-dependent (or magnitude- and time-dependent) weights. One can represent these weights by the quantities $w_1, w_2, \ldots$ (indexed by *k*, the magnitude-bin number; note also that these weights do not necessarily add to 1, and the resulting weighted-likelihood function is

$$\ell(\nu,\beta) = \prod_{k=1}^{\infty} (\nu\ ATp_k)^{n_k w_k} \exp\left[-\nu A \sum_{k=1}^{\infty} Tp_k w_k \right] \qquad (5.3.2\text{-}8)$$

The introduction of weights, resulting in Equation 5.3.2-8, is equivalent to raising the portion of the likelihood function associated with magnitude bin *k* to a power $w_k \le 1$. Raising to a power less then 1 has the effect of flattening this portion of the likelihood, without shifting the location of its peak. For instance, if there are n_1 earthquakes in the first magnitude bin, and this bin is given a weight of 0.1, the associated portion of the likelihood becomes flatter (as if it had only $n_1/10$ earthquakes), but the location of its peak remains unchanged (recall Figures 5.3.2-1 and 5.3.2-2).

We note that the summation in the above equation cannot be simplified after the introduction of the weights. As a result, it is impossible to obtain the maximum-likelihood estimates of ν and β in closed form.

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

$$T_{E,k} = \int_{t_{0,k}}^{present} P_D(t;k)\,dt \qquad (5.3.2\text{-}9)$$

where $P_D(t;k)$ is the detection probability for magnitudes in bin *k* at time *t*. The detection probabilities and equivalent periods of completeness are calculated in Section 3.5 using the EPRI-SOG methodology for determining completeness (which includes a penalized-likelihood analysis of recurrence). For the purposes of this discussion, it is assumed that the equivalent period of completeness is the same within the source zone under consideration and for all magnitudes in a certain magnitude bin. In the final model to be developed here, it will be assumed that the equivalent periods of completeness are constant within each cell comprising the source zone. Another assumption in this derivation is that all earthquakes under consideration (which have magnitudes $m_1, m_2, \ldots m_N$) occurred after their corresponding dates $t_{0,k(m_i)}$.

After introducing magnitude-dependent completeness, the likelihood function becomes

$$\ell(\nu,\beta) = \prod_{k=1}^{\infty} (\nu\ AT_{E,k}p_k)^{n_k w_k} \exp\left[-\nu A\sum_{k=1}^{\infty} T_{E,k}p_k w_k\right] \qquad (5.3.2\text{-}10)$$

A fourth modification is necessary to account for maximum magnitude $M_{\max}$. In the case of a single $M_{\max}$ value, Equation 5.3.2.7 is modified by considering that portions of a magnitude bin, or the entire bin, may be above $M_{\max}$, resulting in the expression

$$p_k(M_{\max}) = \begin{cases} \exp\left[-\beta(m_0^k - m_0)\right] - \exp\left[-\beta(m_0^{k+1} - m_0)\right] & \text{for } m_0^{k+1} < M_{\max} \\ \exp\left[-\beta(m_0^k - m_0)\right] - \exp\left[-\beta(M_{\max} - m_0)\right] & \text{for } m_0^k < M_{\max} < m_0^{k+1} \\ 0 & \text{for } M_{\max} < m_0^k \end{cases} \qquad (5.3.2\text{-}11)$$

There should be a normalizing constant in the above equation, but this constant is very nearly equal to unity (because m_0 is 2.9 and $M_{\max}$ is 5.5 or greater) and can be neglected in this case. For the case where there is epistemic uncertainty in $M_{\max}$ (which is represented in this project by a five-point discrete distribution), one can use the expected value (with respect to $M_{\max}$) of the p_k given by Equation 5.3.2-11; that is,

$$p_k = E_{M_{\max}}\left[p_k(M_{\max})\right] \qquad (5.3.2\text{-}12)$$

where $E_{M_{\max}}[\bullet]$ denotes the expect-value operation with respect to uncertainty in $M_{\max}$. Another consequence of introducing $M_{\max}$ is that the number of magnitude bins becomes finite. In the equations that follow, Q will denote the highest bin with a nonzero value of p_k.

Up to this point, the formulation does not include spatial variation of parameters within a source zone. Consideration of spatial variation is necessary in this project because the assumption of geographically constant rate and *b*-value is not necessarily applicable to some of the large source zones developed in Chapter 4. There are two commonly used approaches to introduce spatial

variation of seismicity within a source zone; namely, the penalized-likelihood (EPRI-SOG, 1986) and kernel approaches (e.g., Frankel, 1995; Stock and Smith, 2002). This project uses the penalized-likelihood approach, with a number of methodological refinements. A discussion of the two approaches and their merits is presented in Section 5.3.2.4.

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

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

$$\ell(\boldsymbol{\nu},\boldsymbol{\beta}) = \prod_{j=1}^{M}\left\{\prod_{k=1,}^{Q}\left(\nu_j A_j T_E(k,j)\,p(k,j)\right)^{n(k,j)w_k}\ \exp\left[-\nu_j A_j \sum_{k=1}^{Q} T_E(k,j)\,p(k,j)\,w_k\right]\right\} \qquad (5.3.2\text{-}13)$$

where vectors $\boldsymbol{\nu},\boldsymbol{\beta}$ contain the values of ν and β in all cells within the source zone, j is the index for the cells (j=1, …, M). Also, the notation has been changed slightly from previous equations, so that $T_E(k,j)$, $n(k,j)$, and $p(k,j)$ are the equivalent period of completeness, earthquake count, and truncated-exponential bin probability (Equation 5.3.2.12) for magnitude bin k and cell k.

Because cells are small and most cells do not contain enough earthquakes to allow the reliable estimation of ν_j and β_j using these earthquakes alone, and because very large variations in ν and β (particularly the latter) between adjacent cells are not considered physically realistic, one introduces penalty functions that penalize those solutions where ν and β have large variations between adjacent cells. Thus the solution to the penalized-likelihood problem represents an optimal compromise between consistency with the data (as indicated by a high value of the likelihood) and smoothness (as indicated by a low value of the penalty function). For a given value of the smoothing parameters, the likelihood has a stronger effect on the solution in cells where data are abundant because only a narrow range of values are consistent with the data.

This project utilizes a penalty function based on the Laplacian operator, which represents a generalization of the second derivative to multiple dimensions, and is commonly used in applications of this kind (e.g., Fukuda and Johnson, 2008). For a function $f(x,y)$ in two dimensions, the Laplacian operator and its lowest-order finite difference approximation are given by

$$\begin{aligned}\Delta f &= \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \approx \\ &\approx 2\left(\frac{f(x-\Delta x, y)/2 - f(x,y) + f(x+\Delta x, y)/2}{\Delta x^2}\right) + \\ &\quad 2\left(\frac{f(x, y-\Delta y)/2 - f(x,y) + f(x, y+\Delta y)/2}{\Delta y^2}\right) \\ &= -2\left(\frac{f(x,y) - \overline{f}^{x}(x,y)}{\Delta x^2}\right) - 2\left(\frac{f(x,y) - \overline{f}^{y}(x,y)}{\Delta y^2}\right)\end{aligned} \qquad (5.3.2\text{-}14)$$

where $\overline{f}^{x}(x,y)$ is the average value of f in the two cells adjacent to the (x,y) cell in the x direction. In the special case of $\Delta x = \Delta y$, the finite-difference Laplacian at (x,y) is proportional to the difference between $\overline{f}^{x}(x,y)$ and the average of the values in the four adjoining cells (not counting diagonal cells that share a vertex with the cell under consideration). This project defines the Laplacian in terms of distances in kilometers (rather than in degrees) because this makes more physical sense, resulting in $\Delta x / \Delta y = \cos(\textit{Latitude})$ and yielding a slightly more complicated expression for the finite-difference Laplacian. Based on the above, and assuming a normal distribution for variations in β between a cell and its neighbors, the penalty function for β over the entire source source zone is of the form

$$f_{\beta}(\boldsymbol{\beta}; \sigma_{\Delta b}) = \prod_{j=1}^{M} \frac{1}{\sqrt{2\pi}\sigma_{\Delta\beta}} \exp\left[-\frac{1}{2}\left(\frac{\Delta\beta_j}{\sigma_{\Delta\beta}}\right)^2\right] \qquad (5.3.2\text{-}15)$$

where the product extends over all cells, $\sigma_{\Delta\beta}$ controls the degree of smoothing, and

$$\Delta\beta_j = -\frac{1}{2}\left\{\left(\frac{\beta(x_j,y_j) - \overline{\beta}^{x}(x_j,y_j)}{\cos^2(\text{Latitude}_j)}\right) + \left(\frac{\beta(x_j,y_j) - \overline{\beta}^{y}(x_j,y_j)}{1}\right)\right\} \qquad (5.3.2\text{-}16)$$

Because Δy is constant, it can be absorbed into $\sigma_{\Delta\beta}$. The change from a factor of 2 to a factor of ½ in the definition of $\Delta\beta_j$ is introduced for the sake of consistency with the EPRI-SOG formulation, and its effect is absorbed into $\sigma_{\Delta\beta}$. For a cell at the edges of the source zone, $\overline{\beta}^{x}(x_j,y_j)$ or $\overline{\beta}^{y}(x_j,y_j)$ is calculated using only cells on one side.

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

It is useful to consider the effect of $\sigma_{\Delta\beta}$ (and the corresponding $\sigma_{\Delta\nu}$ for rate), which control the degree of smoothing. A lower value favors solutions with less variation between adjacent cells. In the limit, infinitesimally small values of $\sigma_{\Delta\nu}$ and/or $\sigma_{\Delta\beta}$ lead to constant ν and/or β throughout the source zone. At the other extreme, very large values of $\sigma_{\Delta\nu}$ and $\sigma_{\Delta\beta}$ is equivalent to determining ν and β separately for each cell, considering only the earthquakes in that cell, without any consideration of spatial continuity.

The two penalty terms, $f_{\nu}(\mathbf{v};\sigma_{\Delta\nu})$ and $f_{\beta}(\boldsymbol{\beta};\sigma_{\Delta\beta})$, multiply the likelihood function in Equation 5.3.2-13, favoring solutions that are consistent with the data and, simultaneously, show less variation between parameters $\ln(\nu)$ and β in adjacent cells. One can fix the penalty terms $\sigma_{\Delta\nu}$ and $\sigma_{\Delta\beta}$ and then solve the penalized-likelihood problem. In this case one can view the penalty terms as prior distributions on the degree of smoothness of $\ln(\nu)$ and β.

Alternatively, one can solve simultaneously for the optimal values of ν, β, and $\sigma_{\Delta\nu}$ and $\sigma_{\Delta\beta}$. In this case, $\sigma_{\Delta\nu}$ and $\sigma_{\Delta\beta}$ are often called hyperparameters. Larger values of these hyperparameters reduce the penalizing effect of the exponential penalty term in Equation 5.3.2-15 and tend to increase the value of the penalized likelihood (by allowing a very close match between model predictions and data in each individual cell). At the same time, these larger values of $\sigma_{\Delta\beta}$ in the denominator of Equation 5.3.2-15 tend to decrease the value of the penalized likelihood. The value of $\sigma_{\Delta\beta}$ at which this optimal balance is achieved depends on the earthquake data, in particular, how many earthquakes there are and how they are distributed in latitude-longitude-magnitude space. This project will refer to the estimates of $\sigma_{\Delta\nu}$ and $\sigma_{\Delta\beta}$ obtained using this approach as the objective estimates of these hyperparameters. They are objective in the sense that they are determined automatically from the data, using the mechanism described above.

This optimal balance between the exponential term and the denominator in Equation 5.3.2-15, which leads to the objective estimate of the smoothing parameter, arises frequently in statistics. Perhaps the best known example is the calculation of the standard deviation of normally distributed data using the method of maximum likelihood.

EPRI-SOG (1986) used a similar formulation for the penalty terms, except that the cosine term is not included, the average of neighboring cells includes diagonal cells that share a vertex with the cell under consideration, and the smoothing parameters are assessed by the expert teams on the basis of their evaluations, aided by statistical tests. The presence of the cosine term is justified on physical arguments (see above); the omission of the diagonal terms makes the definition consistent with the standard finite-difference definition of the Laplacian operator.

Following EPRI-SOG (1986), this project introduces an additional penalty term, in the form of a prior distribution for β_j. The product of the prior distributions for all cells in the source zone takes the form

$$f'_{\beta}(\boldsymbol{\beta}) = \prod_{j=1}^{M} \frac{1}{\sqrt{2\pi}\sigma_{\beta}} \exp\left[-\frac{1}{2}\left(\frac{\beta_j - \beta_{\text{prior}}}{\sigma_{\beta}} \right)^2 \right] \qquad (5.3.2\text{-}17)$$

It may possible to determine β_{prior} and σ_{β} objectively, but this was not done in this project. Instead, β_{prior} was specified as the value of β obtained for the entire study region (see regional *b*-values obtained in Section 3.5), and σ_{β} was selected on the basis of the EPRI-SOG (1986) experience and sensitivity runs.

There is a subtle difference between Equations 5.3.2-15 and 5.3.2-17. The former favors solutions with spatially uniform β, regardless of the value of β, while the latter promotes values of β near β_{prior}.

The complete penalized-likelihood, which may also be viewed as a posterior distribution, is constructed as the product of the joint likelihood for all the cells in the source zone (Equation 5.3.2-13), the smoothness penalty terms (Equation 5.3.2-15 and its counterpart for $\ln(\nu)$), and the prior distribution of β (Equation 5.3.2-17). We write this quantity as

$$p(\mathbf{X}) = \ell(\boldsymbol{\nu}, \boldsymbol{\beta}) f_{\nu}(\boldsymbol{\nu}; \sigma_{\Delta\nu}) f_{\beta}(\boldsymbol{\beta}; \sigma_{\Delta\beta}) f'_{\beta}(\boldsymbol{\beta}) \qquad (5.3.2\text{-}18)$$

where

$$\mathbf{X} = \left[\boldsymbol{\nu}, \boldsymbol{\beta}, \sigma_{\Delta\nu}, \sigma_{\Delta\beta} \right] \qquad (5.3.2\text{-}19)$$

is the complete vector of parameters and hyperparameters. It will be called the state vector in the section that follows.

Equation 5.3.2-18 cannot be solved analytically for the penalized maximum-likelihood estimates of the parameters. Furthermore, we need more than point estimates of the recurrence parameters. Modern PSHA requires an assessment of the epistemic uncertainty associated with these estimates—including correlations between the recurrence parameters of cells in the same geographical region, which may jointly affect the hazard at one site. Therefore, this project needs to characterize the entire multidimensional joint distribution of **X** (this distribution is given by Equation 5.3.2-18 times a normalizing constant). An additional, practical requirement is that one must represent the epistemic uncertainty in a computationally efficient manner. This can be accomplished by means of a small number of alternative "maps" of the recurrence parameters. Sections 5.3.2.1.2 and 5.3.2.1.3 present an approach to fulfill this need.

5.3.2.1.2 Modeling the Joint Distribution of Recurrence Parameters

The approach used to generate alternative maps of the recurrence parameters uses a technique known as Markov Chain Monte Carlo (MCMC), in particular, the Metropolis algorithm (Metropolis et al., 1953). MCMC is frequently used to generate multiple realizations from a complex multi-dimensional probability distribution by constructing a Markov chain that has this distribution as its limiting or stationary distribution. In this case, the distribution of interest is given by Equation 5.3.2-18 (times a constant). This distribution represents the central tendency and epistemic uncertainty of the vector $\mathbf{X} = \left[\boldsymbol{\nu}, \boldsymbol{\beta}, \sigma_{\Delta\nu}, \sigma_{\Delta\beta}\right]$, which contains the recurrence parameters for all cells, plus the two hyperparameters. The number of dimensions of $\mathbf{X}$ is $2M + 2$, where M is the number of cells in the source zone.

A Markov chain is a discrete-time[1] probabilistic model with states $\mathbf{X}^{(1)}$, $\mathbf{X}^{(2)}$, $\mathbf{X}^{(3)}$, $\mathbf{X}^{(4)}$, …., in which the conditional probability distribution of the state at time t+1 (denoted $\mathbf{X}^{(t+1)}$), given the states at earlier times, depends only on the immediately previous state ($\mathbf{X}^{(t)}$). This conditional probability distribution of $\mathbf{X}^{(t+1)}$ given $\mathbf{X}^{(t)}$ is known as the transition probability. If a Markov chain meets certain requirements, it possesses a limiting or stationary distribution, which will be reached asymptotically after many realizations, regardless of the initial state. A mechanical analogy is a linear damped oscillator under harmonic loading. After a sufficiently long time, the oscillator will reach a steady-state amplitude (and, as a consequence, a steady-state probability distribution for the instantaneous displacement), regardless of its initial conditions. Further details on Markov chains may be found in Parzen (1962).

MCMC constructs a discrete-time Markov chain with the following properties: (a) the Markov chain has a stationary or limiting distribution, and (b) this limiting distribution is equal to the joint distribution of interest (i.e., the penalized likelihood function $p(\mathbf{X})$ given by Equation 5.3.2-18). Thus, one can use this sequence to obtain realizations from that distribution. A number of initial realizations are discarded because they are affected by the initial conditions. Then, a number of realizations are generated and summary statistics are calculated from them.

In the Metropolis MCMC algorithm, a new realization $\mathbf{X}^{(t+1)}$ is generated from $\mathbf{X}^{(t)}$ by the following two-step procedure: (1) generate a new candidate state $\mathbf{X}'$ by drawing from a trial distribution $q(\mathbf{X}'|\mathbf{X}^{(t)})$ that depends only on $\mathbf{X}^{(t)}$ and is symmetrical (i.e., $q(\mathbf{X}'|\mathbf{X}^{(t)}) = q(\mathbf{X}^{(t)}|\mathbf{X}')$), and (2) accept the new trial state with probability

$$P_{\text{accept}} = \min\left(1, \frac{p(\mathbf{X}')}{p(\mathbf{X}^{(t)})}\right) \qquad (5.3.2\text{-}20)$$

If the new state is accepted (i.e., if a standard uniform random number generator draws a value lower than P_{accept}), $\mathbf{X}'$ becomes the new state $\mathbf{X}^{(t+1)}$. Otherwise, the new state $\mathbf{X}^{(t+1)}$ is equal to $\mathbf{X}^{(t)}$.

[1] The word "time" is not used in a strict chronological sense here. Instead it is used to represent a sequence of discrete realizations of a mathematical process.

In this project, the Metropolis algorithm is implemented locally, as follows. For each cell *j*, trial values of $\ln(\nu_j)$ and β_j are drawn from uniform distributions centered at the current values of these parameters, the penalized likelihood function is calculated for these trial values, and the values are accepted or rejected using Equation 5.3.2-20 and a random-number generator. The only portions of the penalized likelihood that are reevaluated each time are the likelihood function for cell *j*, the penalty terms for cell *j* and for its neighboring cells within the source zone, and the prior distribution of β_j. After going through all cells, trial values of the hyperparameters $\sigma_{\Delta\nu}$ and $\sigma_{\Delta b}$ are drawn in a similar manner, the penalized likelihood is recalculated, and the values are accepted or rejected. In this step, only the smoothness-penalty terms for all cells are recalculated. This process is repeated a large number of times.

The widths of the uniform distributions are controlled by the analyst. These widths affect the acceptance rate, which in turn affects the numerical efficiency of the algorithm. Recommendations as to the optimal acceptance rates vary; a common recommendation is values in the range between 20% and 40% (e.g., Ntzoufras, 2009). After discarding some of the initial realizations of $\mathbf{X}^{(t)}$, one generates a large number of realizations to represent the joint distribution of the state vector $\mathbf{X} = \left[\boldsymbol{\nu}, \boldsymbol{\beta}, \sigma_{\Delta\nu}, \sigma_{\Delta\beta}\right]$. The section that follows describes the approach to obtain eight equally likely realizations of $\mathbf{X}$ (i.e., eight alternative maps to represent the joint distribution of ν and β and the two hyperparameters), including their correlation structure.

5.3.2.1.3 Development of Alternative Recurrence Maps

The approach presented below constructs alternative maps of the recurrence parameters ν and β that jointly represent the central tendency and statistical uncertainty in the recurrence parameters, including correlations between the recurrence parameters of cells in the same geographical region (which may jointly affect the hazard at one site). At the same time, the number of alternative recurrence maps must be kept small for the sake of efficiency in the hazard calculations. The approach used here is a combination of well known techniques.

The first step in this formulation is to generate many realizations of the state vector $\mathbf{X} = \left[\boldsymbol{\nu}, \boldsymbol{\beta}, \sigma_{\Delta\nu}, \sigma_{\Delta\beta}\right]$ using the MCMC algorithm described above and then use these realizations to calculate the mean vector $\mathbf{m}_{\mathbf{X}}$ and the covariance matrix $\mathbf{S}_{\mathbf{X}}$ of $\mathbf{X}$. This covariance matrix contains information about the marginal variance of ν and β in each cell, the correlation between ν and β in each cell, as well as the correlation between parameters in neighboring cells.

The second step is to perform an eigenvalue decomposition of the covariance matrix $\mathbf{S}_{\mathbf{X}}$. Because this matrix is a covariance matrix, it is positive-semi-definite, so that all eigenvalues are positive or zero. Let $\alpha_1^2, \alpha_2^2, \ldots, \alpha_{2M+2}^2$ be the eigenvalues of $\mathbf{S}_{\mathbf{X}}$, ordered from largest to smallest, and let $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_{2M+2}$ be the associated eigenvectors.

An important property of this eigenvalue decomposition is that of orthogonality, meaning that the dot product of φ_r and φ_s is equal to 1 if $r = s$ and 0 if $r \neq s$. Taking advantage of orthogonality, one can show that a random vector $\widetilde{\mathbf{X}}$ constructed using the following randomized linear combination of the eigenvalues

$$\widetilde{\mathbf{X}} = \mathbf{m}_{\mathbf{X}} + \sum_{k=1}^{2M+2} \varepsilon_k \boldsymbol{\varphi}_k \qquad (5.3.2\text{-}21)$$

where $\varepsilon_1, \varepsilon_2, \ldots \varepsilon_{2M+2}$ are independent random variables with mean zero and variances $\alpha_1^2, \alpha_2^2, \ldots, \alpha_{2M+2}^2$, has mean $\mathbf{m}_{\mathbf{X}}$ and covariance matrix $\mathbf{S}_{\mathbf{X}}$. This implies that a synthesized realization of $\widetilde{\mathbf{X}}$ generated using Equation 5.3.2-21 has the correlation properties of the random vector $\mathbf{X} = [\mathbf{v}, \boldsymbol{\beta}, \sigma_{\Delta v}, \sigma_{\Delta\beta}]$ that we want to simulate. This technique, based on the eigenvalue decomposition of the covariance matrix, is used in many disciplines, under a variety of names; e.g., Karhunen–Loève expansion, principal component analysis, and empirical orthogonal functions (see, for example, Preisendorfer, 1988).

In many applications of this technique, only a few terms are required in Equation 5.3.2-21 because the eigenvalues decrease rapidly in size. This is not the case in this application, and it may be necessary to use a large number of terms (up to a few hundred) in the summation in order to represent 99% of the total variance of **X**.

The third step is to generate eight realizations of the ε's and then use these to generate realizations of $\widetilde{\mathbf{X}}$ using Equation 5.2.2-21, with each realization representing an alternative recurrence map. This is done using a technique known as Latin Hypercube sampling or LHS (Iman and Conover, 1980). The technique will be described in the context of Equation 5.3.2-21, where we want to generate eight realizations of the random vector $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \ldots \varepsilon_{2M+2}]$ and we want these realizations to provide a good representation of the underlying distributions, despite the small number of realizations. We also assume, for the sake of convenience, that the components of $\boldsymbol{\varepsilon}$ have normal distributions (this assumption and its implications are discussed later). The process is as follows:

1. Divide the sample space of each normally distributed ε_j into eight intervals, with each interval containing an equal probability of 1/8. These intervals will have unequal lengths.
2. Within each interval *z* (*z*=1, 2, „,8), generate a random value ε_{jz} that follows the conditional distribution within the interval. The resulting values form a sequence with $\varepsilon_{j1} < \varepsilon_{j2} < \ldots < \varepsilon_{j8}$.
3. For each *j,* generate an independent random permutation of $\varepsilon_{j1}, \varepsilon_{j2}, \ldots, \varepsilon_{j8}$. After this permutation, $\varepsilon_{j1}, \varepsilon_{j2}, \ldots, \varepsilon_{j8}$ are no longer an increasing sequence. The result is eight random realizations of the random vector $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \ldots \varepsilon_{2M+2}]$. Because the eight values for each *j* were drawn from the eight equal-probability intervals, the distribution of ε_j is well

approximated by the eight realizations. Because of the random permutations, the correlation coefficient between the eight realizations of any two ε_{j_1} and ε_{j_2} has an expected value of 0 and is likely to be low.

4. Generate eight alternative realizations of $\widetilde{\mathbf{X}}$ using the eight permuted $\boldsymbol{\varepsilon}$ vectors and Equation 5.3.2-21.

5. Convert each realization of $\widetilde{\mathbf{X}}$ (actually, all but its last two elements) into values of ν and b for each cell in the source zone, and use these two values to calculate $\nu(5)$, the rate of earthquakes above magnitude 5.

The resulting eight alternative maps represent the uncertainty in recurrence parameters that results from the limited duration of the catalog. If the smoothing parameters are treated as uncertain and estimated objectively from the data, the eight alternative maps also include the uncertainty about the appropriate values of the smoothing parameters.

This approach presented below constructs the alternative realizations of ν and β as linear combinations of normally distributed random quantities. Thus, it produces normally distributed quantities. This normal distribution is consistent with the distribution shapes of ν and β shown in Figures 5.3.2-1 and 5.3.2-2, which approach normal shapes as the number of earthquakes approaches 10. Although there may be a few earthquakes in each cell, the presence of the penalty terms tends to "pool" the data from many adjacent cells, resulting in a larger value of the "effective" earthquake count.

The initial implementation of this approach required that the number of realizations be a power of 2 because the first few epsilons were sampled using two-point distributions. This number was set to 8 because 4 was considered insufficient and 16 imposed a high computational burden for the hazard calculations. In the present Latin Hypercube implementation, the restriction of a power of 2 no longer exists, but the choice of eight realizations was retained. Tests indicate that eight realizations, together with Latin Hypercube sampling, provide an adequate representation of the mean and fractiles of the hazard.

5.3.2.2 Application of the Model and Specification of Model Parameters

5.3.2.2.1 Calculations

Calculations were performed for all the seismotectonic and Mmax-region source zones identified in Chapter 4. The catalog had all dependent earthquakes removed; earthquakes associated with the RLME sources were also removed from this catalog to avoid double-counting (see Chapter 3).

The MCMC formulation was used to generate a large number of realizations (typically 15 million), of which roughly the first 25% were discarded because they were affected by the initial conditions. The evolution of recurrence parameters at a few points was examined in each run to verify that this number was sufficient, in the sense that the system had reached a steady state. In some instances, this number had to be increased and sometimes the width of the MCMC trial

distribution had to be altered. The cell sizes for all seismotectonic source zones except MidC was 0.25 degrees in order to have the ability to resolve any sharp gradients in parameters that may occur in the more active regions. For reasons of computational efficiency, the cell size for MidC and for the Mmax source zones was set to 0.5 degrees. Tests with MidC indicated similar results for both cell sizes.

In addition to the algorithmic inputs described above, one must specify the weights for the various magnitude bins, the degree of spatial smoothing (including the option of objective smoothing), and the mean value and standard deviation of the prior distribution on b. Section 5.3.2.1.1 provides details on how these quantities enter the calculations and the rationale for their use.

Table 5.3.2.1 shows the five cases that were initially considered for the weights to the magnitude bins. These five alternative sets of weights span a wide range, in order to investigate the effect of giving lower weights to the lower-magnitude bins. Case E was introduced later, as an intermediate case between Cases C and D, because Case E was considered too extreme.

Some or the reasons for giving lower weights to the lower magnitude bins were discussed earlier. For instance, the magnitude-recurrence law may deviate from exponential, or the magnitude-conversion models or completeness model may be less reliable for lower magnitudes. In this regard, it is useful to remember that only earthquakes with magnitudes greater than 5 (4 if the CAV filter is being used) are important for seismic hazard and risk for nuclear facilities. The only reason for considering lower magnitudes is that the $\mathbf{M} > 4$ data alone are not sufficient for determining both the magnitude-recurrence law and the spatial distribution for earthquakes of engineering interest. Because the magnitude weights also affect the degree of smoothing[2], one must also consider the issue of spatial stationarity; i.e., whether the spatial pattern of past, small, earthquakes is representative of the spatial pattern of future, hazard-significant, earthquakes. Cases A through E were designed to cover a broad range of alternatives regarding magnitude weights.

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

The mean value for the prior distribution of β (Equation 5.3.2.17) was calculated separately for Cases A, B, and E above, by performing a maximum-likelihood recurrence analysis for the entire study region, and assuming constant b. Details of these calculations are contained in Section 3.5

[2] When objective smoothing is used, the cases where the lower bins are downweighted result in smoother seismicity. This effect will be discussed and explained later in this section.

Two values were considered for the strength of the prior distribution of β, which is specified by the standard deviation $\sigma_b = \sigma_\beta / \ln(10)$. The first value is 0.6 (in units of *b*), which corresponds to a "moderate prior" in the EPRI-SOG (1988) project; we will also refer to this value as a moderate prior for b. The second value is 0.3 (in units of *b*), and we will refer to it as a strong prior. We also considered a weak prior (1.2, in units of *b*) in earlier stages of the project, but this value was not considered credible because there is a general consensus in the TI Team that spatial variation in *b* should be mild.

For all source zones except SLR, the smoothing parameters were determined using the objective approach. For each source zone, and for each of the magnitude-bin weights and strengths of the prior on b considered, the smoothing parameters $\sigma_{\Delta\nu}$ and $\sigma_{\Delta\beta}$ were allowed to vary, allowing the catalog data themselves to determine their optimal values and associated uncertainty for each source zone in an objective manner. Figures 5.3.2-4 through 5.3.2-9 show the values of the smoothing parameters and their $\pm$ one-sigma ranges for magnitude-weight Cases A, B, and E, using a moderate prior for b.

On Figures 5.3.2-4 through 5.3.2-9, the source zones are sorted by size, with the small seismotectonic zones on the left side, and the Midcontinent and the Mmax zones on the right-hand side. The EPRI-SOG equivalent values (adjusted for differences in cell size) are provided for the sake of reference; the equivalence is only approximate because of some of the differences mentioned earlier between the two formulations of smoothing and because of potential limitations in the adjustment for size.

On Figure 5.3.2-4, the small- and moderate-size seismotectonic zones show a broad variation in the objectively determined values of $\sigma_{\Delta\nu}$: zones with high seismicity such as Reelfoot Rift show higher values (less smoothing), while zones with low seismicity such as AHEX show lower values (more smoothing). The large MidC and Mmax zones show similar values as the active zones. Relative to the EPRI-SOG values, the objectively determined values indicate less smoothing, except for the seismotectonic zones with low seismicity. Comparing Figures 5.3.2-6 and 5.3.2-8, to Figure 5.3.2-4 we note that the objective approach assigns significantly more smoothing to the cases where the lower-magnitude bins are downweighted. The reason for this is that the effective number of earthquakes in a source-zone is diminished when the low-magnitude data are down-weighted. The objective approach compensates for this smaller sample size by imposing more smoothing, thereby reducing spatial resolution.

On Figure 5.3.2-5, there is zone-to-zone variation in the objectively determined values of $\sigma_{\Delta b}$, there is a trend with size, but no trend with seismicity. It is important to notice that there is significantly more smoothing on *b* than for EPRI-SOG, indicating that the data favor a very mild geographical variation in *b*. Comparing Figure 5.3.2-7 and 5.3.2-9 to Figure 5.3.2-3, we note that the objective approach assigns slightly higher smoothing on b for the cases in which the low-magnitude bins are downweighted.

Figures 5.3.2-10 and 5.3.2-11 illustrate the effect of smoothing and seismic activity by comparing the results for the extended-crust source zones ECC-AM and ECC-GC, for case-A magnitude weights and moderate prior on *b*. ECC-AM has higher activity and the objective

procedure assigns to it low smoothing for rates (high $\sigma_{\Delta\nu}$). In contrast, ECC-GC has low activity and the objective procedure assigns to it approximately 10 times more smoothing for rates. Note that the scales for activity rate are different in the two figures.

It is also useful to compare the above figures to Figures 5.3.2-12 through 5.3.2-15, which show results for the same source zones, for Case B and E magnitude weights and moderate prior on *b*. Because the effective number of earthquakes is smaller, the objective procedure compensates by increasing the smoothing on rate. The effect is much more pronounced in ECC-AM.

Figures 5.3.2-10 through 5.3.2-15 show that, by considering three alternative sets of magnitude weights together with an objective approach for smoothing, we are effectively sampling a broad range of smoothing parameters for rate. To further appreciate this, it is useful to examine the mean recurrence maps for Cases A, B, and E that will be presented in Sections 6.4 and 7.5.

For the SLR source zone, the approach described above led to underestimation of the observed rates for magnitude 4.3 and higher. To obtain an adequate fit it was necessary to remove the prior distribution on β and to constrain the smoothing on β. This constraint was introduced by specifying a prior distribution of $\sigma_{\Delta\beta}$ with a median of 0.01 and a COV of 30%, which appears as an additional multiplicative term in Equation 5.3.2-18. The need for this different treatment is that SLR appears to have a *b*-value that is substantially lower than that of the region as a whole, and that portions of SLR (particularly in the Charlevoix area) appear have even lower *b*-values. This is consistent with the experience of Petersen et al. (2008), who assign a lower *b*-value to Charlevoix. Figures 5.3.2-10 through 5.3.2-15, and all other figures shown later in this section and in Chapter 7 reflect this modified approach for the treatment of the SLR source zone.

5.3.2.2.2 Sensitivity to Magnitude Weights, Strength of Prior, and Alternative Maps

The six sets of recurrence maps obtained above were used to perform sensitivity calculations in order to quantify and illustrate the effect of the strength of the prior on *b* and magnitude weights, and alternative recurrence maps. Calculations were performed for the host zones under the most likely source-zone configuration. These calculations were also performed at other demonstration sites, but only representative results will be shown here.

Figures 5.3.2-16 and 5.3.2-17 show the sensitivity to the strength of the prior on *b*, for the Manchester and Topeka sites. These results show essentially no effect on hazard, a result that was observed at other demonstration sites and for both PGA and 1 Hz SA. As a consequence of these comparisons, the strong prior on *b* is deemed unnecessary and is eliminated from further consideration.

Figures 5.3.2-18 and 5.3.2-19 show the sensitivity to magnitude weights for the same two sites. These results indicate a moderate effect on hazard, particularly in areas of low seismicity. Finally, Figures 5.3.2-20 and 5.3.2-21 show the effect of the eight alternative recurrence maps on the hazard at the same two sites, for Case B magnitude weights. In both cases, the effect is important.

Based on the above comparisons, only the moderate prior on *b* will be considered (except for SLR, where no prior is used), and the logic-tree branches associated with magnitude weights and the eight alternative realizations will be retained.

5.3.2.2.3 Resulting Maps

Figure 5.3.2-22 shows the map of mean recurrence parameters for the most-likely source-zone map in the logic tree. Figure 5.3.2-23 shows the associated uncertainties, in the form of COV of the rate and standard deviation of the *b*-value. Figures 5.3.2-24 and 5.3.2-25 show two realizations of the equally likely recurrence maps. Although the last two maps show areas of no activity (i.e., white areas), it is important to note that no region has a mean rate of zero (i.e., these white areas "move" from one map to another), and that these white areas almost always occur in areas of no seismicity. Note also that the areas of high activity change little between the mean map and the two realizations, indicating that the uncertainty in rate (in the form of the COV) is low in regions with abundant data.

Maps of mean recurrence parameters for all source-zone configurations are shown in Sections 6.4 (for Mmax zones) and 7.5 (for seismotectonic zones). All realizations of the alternative maps are shown in Appendix J.

5.3.2.3 Exploration of Model Results in Parameter Space

This section explores the recurrence parameters obtained in the previous section by examining and discussing the results obtained at a few specific locations and comparing the recurrence parameters to the catalog data. The purpose of these comparisons is to demonstrate consistency with the data under a variety of conditions and to gain insights into the behavior of the model developed here. These comparisons will focus on the higher-weighted branches of the master logic tree (namely, the seismotectonic zonation with the narrow interpretation of PEZ and Rough Creek graben in the Midcontinent). Figure 5.3.2-26 shows the local areas considered in these comparisons.

5.3.2.3.1 Regions of High Activity

The first region to consider is East Tennessee, which is represented by the same polygon used by Petersen et al. (2008; see Figure 5.3.2-26). The purpose of this comparison is to demonstrate that the penalized-likelihood formulation does not cause unwarranted dilution of highly localized seismicity.

Figures 5.3.2-27 through 5.3.2-29 compare the observed seismicity within the polygon to the expected counts within the same polygon, as calculated from the recurrence maps source zone PEZ-N, which is the host zone for this polygon. For each recurrence map, the expected counts are calculated by (1) calculating the expected rate for each magnitude bin using the values of ν and β for the cells covered by the polygon, taking into account that some cells are only partially contained within the polygon and accounting for $M_{\max}$ (recall Equations 5.3.2-11 and 5.3.2-12),

(2) multiplying this rate by the equivalent period of completeness for that bin, and (3) summing the expected counts, beginning at the highest bin. The observed counts do not take integer values because they represent effective counts. The error bars on the data represent the 16%–84% range of statistical uncertainty and are calculated using the approach by Weichert (1980). Each of the eight curves represents expected counts (i.e., rate times equivalent period of completeness) for one of the eight equally weighted alternative maps. Figures 5.3.2-27 through 5.3.2-29 use the maps for magnitude-weight Cases A, B, and E, respectively.

Figures 5.3.2-27 and 5.3.2-28 show a close agreement between model and data for Cases A and B. As anticipated, the results for Case A show a better fit to the low-magnitude data, while the results for Case B show a better fit for magnitude 4.3 and higher. Both figures also show that the eight curves are closely spaced, indicating low uncertainty in recurrence parameters in areas of high activity.

The results for Case E on Figure 5.3.2-29 shows lower rates, with the expected counts slightly below the error bars for the **M** 4.3–5.0 magnitude bin. Because the objective smoothing imposes more smoothing for Case E, it is not surprising that the expected counts within a small region of high activity are somewhat lower than the observed counts. Nonetheless, the expected counts from Cases A, B, and E—taken as an ensemble—are consistent with the data, especially when the error bars are taken into account.

Figures 5.3.2-30 and 5.3.2-32 show similar results for a $3° \times 2°$ rectangular region in central New England, comprising Massachusetts and southern Vermont and New Hampshire (see Figure 5.3.2-26). This region overlaps the NAP and ECC-AM seismotectonic zones, which are used to calculate the expected counts. Again, there is good agreement between model and data and the uncertainty in recurrence is low.

5.3.2.3.2 Nemaha Ridge Area

The Nemaha Ridge in Kansas and Oklahoma has experienced two **M** > 5.0 earthquakes in historical times, despite being in an area of moderate seismicity in the Midwest. The geometry considered (see Figure 5.3.2-26) is a rough average of the geometries considered by the EPRI-SOG (1988) project for the Nemaha Ridge.

Figures 5.3.2-33 through 5.3.2-35 compare seismicity and recurrence parameters in the same manner used above. The two earthquakes in the **M** 5.0–5.7 interval (the highest intervals with data) are the April 24, 1867, and April 9, 1952, earthquakes. Figures 5.3.2-33 and 5.3.2-34 (Cases A and B) show predicted rates that are at or slightly below the one-sigma error bars, indicating marginal agreement with the data. Figure 5.3.2-35 (Case E), shows less agreement. The explanation for this result is that the effect of giving lower weights to the first two magnitude bins—which one would have expected to produce a better fit for the higher bins—is counteracted by higher smoothing.

Although the rates of **M** > 5 earthquakes for this region are slightly underestimated, it is important to keep in mind that the TI team, after reviewing the literature on the Nemaha Ridge,

did not see a justification for identifying the Nemaha Ridge as a distinct seismotectonic zone. Two **M** > 5 earthquakes in this region are not sufficient to change that assessment.

5.3.2.3.3 Regions of Low Activity

Kernel methods have a tendency to predict very low rates in regions with few or no recorded earthquakes because the commonly used Gaussian kernel function decays very rapidly at distances greater than twice the kernel width. This problem is discussed by Frankel et al. (1996), where the problem is avoided by introducing an alternative "floor" based on a constant weight.

The penalized-likelihood approach developed here does not have this problem. As indicated in Section 5.3.2.1.2 and illustrated on Figure 5.3.2-1 for the simple formulation considered at first, the likelihood function for zero observed earthquakes allows for nonzero rates and has a mean rate of $1/AT$, which acts as a natural floor for the activity rate. This result also applies to one cell, if there is no penalty function linking the cell's parameters with its neighbors'. Introduction of the penalty function has two effects on this floor. On the one hand, it increases the effective area in the denominator, thereby reducing the mean rate. On the other hand, it also elevates the rate if there is activity near the cell under consideration. Examination of the maps on Figure 5.3.2-16 and those in Sections 6.4 and 7.5 suggests that the rates do not drop rapidly as one moves into regions of very low activity, as happens with the Gaussian kernel (where the rate decays as $\exp[-1/2\,(x/\sigma)^2]$, where x is the distance to the closest earthquake and σ is the kernel size).

Figures 5.3.2-36 through 5.3.2-38 compare seismicity and recurrence parameters for a $3° \times 2°$ rectangular region in the southern Florida Peninsula near Miami (see Figure 5.3.2-26). There is good agreement between the recurrence curves and the one earthquake in the catalog. The mean annual rate of **M** 5 or greater is approximately 0.01 earthquakes (for an area of approximately 5.5 squared equatorial degrees). These figures also show a large spread between the alternative recurrence curves (especially for the reduced-magnitude weights), indicating a high uncertainty in rates.

Figures 5.3.2-39 through 5.3.2-41 provide similar comparisons for a $3° \times 2°$ rectangular region in northern Minnesota, around St. Paul (see Figure 5.3.2-26). Although there are no earthquakes within this area, the model predicts activity rates that are comparable to those obtained for Miami.

These two sets of results demonstrate the effect of a natural floor that arises from consideration of the full likelihood function for the rate. This natural floor has the advantage that it does not require any external assumptions regarding the spatial extent that should be used to determine the floor rate.

5.3.2.4 Consideration of Constant *b*-Value Kernel Approaches

In addition to the penalized-likelihood approach, this project evaluated kernel approaches as alternatives for the calculation of recurrence parameters. Both fixed size (e.g., Frankel, 1995) and adaptive (e.g., Stock and Smith, 2002) were evaluated and exercised in the early stages of this

project. In the end, only the penalized-likelihood approach was retained in the logic tree. The rationale for this decision is provided below.

The first consideration is the specification of kernel size. In most applications of the kernel method for PSHA applications (e.g., Petersen et al., 2008), the kernel size is fixed spatially and is assessed subjectively, although methods exist for the objective selection of kernel size (e.g., Silverman, 1986). The adaptive kernel approach of Stock and Smith (2002) is partially objective. It begins with an arbitrary kernel size, and then it applies a location-dependent adjustment that depends on the density of data. This approach can be made fully objective, as illustrated below. In contrast to the approach used in most previous PSHA applications, the penalized-likelihood approach employed in this project is objective, in the sense that it determines the optimal degree of smoothing for each source zone (recall Figures 5.3.2-4 through 5.3.2-9), finding the best tradeoff between statistical precision and spatial resolution. This consideration of objectivity is not as important as those that follow, however. Although objective approaches are desirable, there are situations in PSHA where informed judgment is required because the data and statistical models are not sufficient.

The second consideration is adaptability. Fixed kernel approaches cannot provide higher spatial resolution where data are abundant and lower resolution where data are sparse. The adaptive kernel approach provides adaptability by performing an initial calculation of rate and then adjusting the kernel size for each location based on the result of this calculation. The penalized-likelihood approach provides adaptability because the likelihood function is narrower where data are abundant and broader where data are sparse (recall Figures 5.3.2-1 and 5.3.2-2). As a result, the relative effect of the penalty terms (which tends to reduce roughness in the recurrence parameters) is weaker where data are abundant and stronger where data are sparse. The comparisons shown earlier for East Tennessee demonstrate the adaptability of the penalized-likelihood approach.

The third consideration is that of geographical variation of *b*. The kernel approaches (both fixed and adaptive), as they have been implemented so far for PSHA, do not allow for geographical variation in *b*. In contrast, the penalized-likelihood approach allows for geographical variation in *b*. Furthermore, smoothing on rate and on *b* is controlled by separate smoothing parameters, allowing for more variation in rate and less variation in *b*, as the data and the technical community's intuition suggest. This ability to allow variation in *b* is important for large source zones thousands of kilometers across and encompassing different tectonic settings, such as those in the Mmax branches of the master logic tree. For these large source zones, some geographical variation in *b* should not be precluded a priori.

The fourth consideration is the quantification of epistemic uncertainty. The two kernel approaches, as they have been implemented so far for PSHA, provide a simplistic representation of epistemic uncertainty in rate and *b*. This representation essentially treats the source zone as having constant rate and *b* for the purposes of calculating uncertainty, and then assigns the resulting COV in rate and standard deviation in *b* to the entire source zone. This approach underestimates the true uncertainty. It is possible to obtain more realistic estimates of the uncertainty by using bootstrapping (i.e., fitting the kernel model to multiple synthetic catalogs), but the number of alternative maps required to obtain a stable estimate of the uncertainty in

hazard is likely to be large. The penalized-likelihood approach combined with MCMC provides a realistic representation of uncertainty in the hazard because it samples the entire penalized-likelihood function for the cells, thereby providing an estimate of the uncertainty in both rate and *b* for each cell, and the associated correlations. In particular, this uncertainty, as measured by the COV, is higher in regions of low activity than in regions of high activity (again, recall Figures 5.3.2-1 and 5.3.2-2). In addition, the penalized-likelihood approach also incorporates the epistemic uncertainty in the smoothing parameters, as represented by the error bars on Figures 5.3.2-4 through 5.3.2-9). Considering multiple kernel sizes in the logic tree (e.g., Petersen et al., 2008) also captures this uncertainty, but in a subjective manner. Furthermore, the penalized-likelihood approach combined with MCMC and with the eigenvalue-based sampling scheme presented here provides an efficient representation of epistemic uncertainty (including correlations between cells) by means of eight alternative maps.

A fifth consideration is the rates in regions of very low seismicity. The fixed kernel approach, and to a lesser extent the adaptive kernel approach, yield very low estimates in these regions. This problem may be resolved by the introduction of a somewhat artificial floor level, or by using a different kernel shape. The penalized-likelihood approach, on the other hand, provides a natural floor that is consistent with the data (or lack thereof) in these regions. This floor is based on the understanding that the observation of zero earthquakes during a certain time interval does not necessarily imply a rate of zero, but it does imply that the rate is unlikely to be moderate or high.

In relation to these regions of low seismicity, there may be valid economical and structural-engineering reasons for imposing minimum design values in building codes—and these values may exceed the hazard computed from the natural or artificial floor values discussed here. These considerations are the responsibility of code-writing committees and should be kept separate from the process of estimating recurrence parameters and computing seismic hazard.

A sixth consideration is that the choice between penalized likelihood and kernel smoothing is a choice between statistical tools, not a choice between different conceptual models regarding the spatial distribution of future seismicity. Both approaches implement the same conceptual model, but apply different statistical tools—each with its own strengths and weaknesses, as discussed above.

To illustrate this last point, it is useful to compare the result from the penalized-likelihood and kernel-smoothing methods for selected source zones (and magnitude-weight Case A), using the same catalog and completeness model. Figure 5.3.2-42 shows the recurrence parameters calculated for three of the seismotectonic zones, spanning a broad range of source-zone size and rate density). These results were obtained using an initial kernel size calculated by cross-validation (Silverman, 1986), and then applying an adaptive-kernel (Stock and Smith, 2002). This approach is an objective and adaptive implementation of the Gaussian kernel method. Comparison of the rates on Figure 5.3.2-42 to the corresponding portions of Figure 5.3.2-22 shows a good agreement in most regions, particularly in regions of high activity such as portions of NAP and regions of moderately low activity such as portions of ECC-AM. The main difference is in the areas of very low activity (e.g., central Texas, northern North Dakota, offshore areas in the northeast of the study region), where the kernel approach predicts

essentially zero rates but the penalized-likelihood approach predicts low (but nonzero) rates. Another fundamental difference is that the kernel approach, as currently implemented, does not produce spatially variable estimates of uncertainty such as those shown on Figure 5.3.2-23.

All these considerations led the project TI Team to the decision to retain only the penalized-likelihood approach (together with MCMC and eigenvalue-based sampling) for the purpose of calculating recurrence parameters and their uncertainty.

5.3.2.5 Comparison to EPRI-SOG Approach

The penalized-likelihood approach used in this project has a number of elements in common with the EPRI-SOG (1988) methodology, but there are also substantial differences in implementation. This section summarizes and discusses the common elements and the differences in the calculation of recurrence parameters between this project and the EPRI-SOG project.

The basic formulation of the problem is the same for this project and the EPRI-SOG project. Both projects divide a source zone into rectangular cells, (using partial cells at the edges to honor the geometry of the source zone), both projects formulate the likelihood function for ν and β in each cell using magnitude bins, both projects introduce penalty functions (in terms of $\ln[\nu]$ and β) that increase continuity between cells, and both projects introduce a prior distribution for β.

There are some differences in the way the penalty terms are computed. In particular, this project uses the standard form of the finite-difference Laplacian operator, using a slight adjustment to correct for the effect of latitude on the East-West size of the finite-difference grid. In contrast, the EPRI-SOG project uses the difference between a cell and the average of all its neighbors (which is not too different from the average of the Lapacians computed along the longitude-latitude grid and along a grid rotated 45 degrees from it, without correcting in either case for difference in the size of the finite-difference grid). The effect of these differences is believed to be small, and the approach used in this project is preferable because it is isotropic with respect to distances in kilometers. There are also differences in parameterization (i.e., this project uses standard deviations $\sigma_{\Delta\nu}$ and $\sigma_{\Delta\beta}$, while the EPRI-SOG project uses quantities that multiply the squared difference between a cell and its neighbors), but these differences are of no consequence.

The main difference between this project and the EPRI-SOG project is in the treatment of epistemic uncertainty. This project captures epistemic uncertainty by working with the full penalized likelihood function, which contains all information about the "center, body, and range" of the distribution of recurrence parameters, as well as information about the correlation between recurrence parameters in nearby cells. This penalized likelihood function is sampled using Markov Chain Monte Carlo (MCMC), which generates many realizations of the recurrence parameters. These realizations are then used to construct eight representative realizations by computing the covariance matrix of the recurrence parameters, and then using a Karhunen–Loève expansion of this covariance matrix, together with Latin Hypercubes to generate a set of eight realizations that represent the mean and epistemic uncertainty in recurrence parameters. In contrast, the EPRI-SOG methodology works only with the maximum penalized likelihood

estimates of the recurrence parameters. The EPRI-SOG study implemented a "bootstraping" procedure in its software (i.e., generating multiple synthetic catalogs and performing the maximum penalized likelihood calculations for each catalog) to quantify uncertainty, but this procedure was never used by the EPRI-SOG (1988) project in production calculations that went all the way to hazard (and it is believed to be impossible in practice, due to the number of alternative recurrence maps required).

Another difference relates to the smoothing parameters $\sigma_{\Delta\nu}$ and $\sigma_{\Delta\beta}$. This project calculated optimal values of these parameters using an objective approach, while the EPRI-SOG project specified multiple values of the equivalent smoothing parameters using expert opinion (aided by statistical diagnostics). This project also treated $\sigma_{\Delta\nu}$ and $\sigma_{\Delta\beta}$ as free hyperparameters in the MCMC, thereby capturing their epistemic uncertainty and the effect of that uncertainty on the uncertainty in $\boldsymbol{\nu}$ and $\boldsymbol{\beta}$.

There is also a small difference in the treatment of magnitude uncertainty in the catalog. This project counts earthquakes by using "event factors" different from unity, as discussed earlier. In contrast, the EPRI-SOG project uses a uniform magnitude M* different from the best-estimate magnitude. Analysis and numerical tests in Chapter 3 indicates that the two approaches are equivalent in principle, but the use of event factors is preferable because it avoids edge effects at the lower magnitudes.

An area in common is the calculation of completeness times. Prior to performing recurrence calculations for each team, source zone, and smoothing option, the EPRI-SOG (1988) project performed a coupled recurrence and completeness analysis, using a zoneless penalized maximum-likelihood formulation for recurrence, together with certain assumptions regarding completeness (e.g., the detection probability does not decrease with time or with magnitude). This project followed the same approach, using the same EPRI-SOG formulation and software, but using the new catalog and an updated map of the completeness region. These calculations are documented in Section 3.5.

5.3.2.6 Assessment of the Lombardi Study

In a study of the 1990–2001 $\mathbf{M} > 1.9$ Southern California earthquakes, Lombardi (2003) observes that the catalog of all events (i.e., main shocks and aftershocks combined) follows an exponential magnitude distribution, but the catalog of main shocks deviate from exponential, with flattening at low magnitudes. Lombardi then proposes the following model: (a) the magnitude distribution for the population of all events is exponential, and (b) the magnitude distribution of main events may be constructed as the distribution of the maximum of N independent identically distributed (iid) exponential magnitudes, where N is the number of events in a cluster (as determined by the declustering algorithm). This model predicts a main-shock magnitude-recurrence curve that deviates from exponential in the M 1.9–2.5 range, and this deviation is observed in the data.

One physical problem with this model is the assumption of iid magnitudes within a cluster. One would expect that the same physical factors that introduce tight spatial and temporal clustering of earthquakes would also introduce correlation in their magnitudes.

There are also a number of alternative explanations for the observed deviation from exponentially distributed main shocks. One possible explanation is that the 1900–2001 Southern California catalog may not be complete down to **M** 1.9. In fact, Felzer (2008) finds that the lowest complete magnitudes for Southern California during the period 1993–1997 are in the **M** 2.7–3.0 range and that completeness down to magnitudes near 2.0 was only achieved during the period 2000–2007. Another possible explanation is edge effects associated with incomplete clusters that include some aftershocks with $\mathbf{M} < 1.9$ and a main event with $\mathbf{M} > 1.9$ (and for which N has been underestimated).

Even if the observed deviation from exponentiality were real and not an artifact of incompleteness or edge effects, there are a number of reasons why it would not have large practical implications for this project. The reason for this is that this deviation is localized, extending only over the lowest 0.3- to 0.5-magnitude units. This project (and, to a lesser degree, EPRI-SOG) use alternative sets of magnitude weights to allow for deviations from exponentiality and other problems that may arise with magnitudes near the cutoff. Recall that magnitude-weight Cases B and E (which, together, carry 70% of the total probability in the logic tree) give 10% and 0% weight to the first magnitude bin (see Table 5.3.2-1). In addition, the formulation of the likelihood function in terms of magnitude bins does not work with the local slope of the magnitude-recurrence data, reducing the effect of any localized deviations from exponentiality. Finally, the completeness analysis for magnitude-weight Cases A and B would likely compensate for any deviation from exponentiality by altering the completeness times.

In summary, even if the deviations from main shock exponentiality postulated by Lombardi (2003) were real, they would have no effect on the *b*-values computed in this project.

5.3.3 Estimation of Recurrence for RLME Sources

The earthquake recurrence models for the RLME sources are somewhat simpler than those described in Section 5.3.2 in that the magnitude range for individual RLMEs is relatively narrow and their spatial distribution is limited geographically such that spatial variability is not an issue. (Note that for simplicity in this section and later in the report, the term *RLM*E is used to refer to the actual past earthquakes and the forecast future occurrences; the term *RLME source* is used to refer to the seismic source used to model the spatial and temporal distribution of the RLMEs.) This limits the problem to one of estimating the occurrence rate in time of a point process. The data that is used to assess the occurrence rates are derived primarily from paleoseismic studies and consist of two types. In many cases, the data provide estimated ages of the paleoearthquakes such that the times between earthquakes can be estimated. In other cases, the data provide an estimate of the number of earthquakes that have occurred post the age of a particular stratigraphic horizon. These data can be easily used to derive estimates of the RLME occurrence rates and their uncertainty.

The estimation of the RLME occurrence rates is dependent on the probability model assumed for the temporal occurrence of these earthquakes. The standard model applied in most PSHA studies and the one adopted for most RLME sources in this study is the Poisson model in which the probability of occurrence of an RLME in a specified time period is completely characterized by a single parameter, λ, the rate of RLME occurrence. The approaches used to assess λ and its uncertainty are described in Section 5.3.3.1. The principal features of the Poisson process are that it is "memoryless"—the probability of occurrence in the next time interval is independent of when the most recent earthquake occurred, and the time between earthquakes is exponentially distributed with a standard deviation equal to the mean time between earthquakes.

For two RLME sources, the data are sufficient to suggest that the occurrence of RLMEs is more periodic in nature (the standard deviation is less than the mean time between earthquakes). For these RLME sources a simple renewal model can also be used to assess the probability of earthquake occurrence. This model takes into account the time that has elapsed since the most recent RLME occurrence in making an estimate of the probability of occurrence in the future. The probability of occurrence for renewal models is time-dependent in that the probability of occurrence in the next time interval is dependent upon the elapsed time since the most recent earthquake. The approaches used to assess the RLME occurrence probabilities for a renewal model are presented in Section 5.3.3.2.

5.3.3.1 Estimation of Occurrence Rates for the Poisson Model

The estimation of the uncertainty distribution for the rate parameter λ of a Poisson process uses the following general framework:

- Define a likelihood function that represents the probability of seeing the observed sample of data given a specified rate parameter λ.
- Compute the likelihood values for a full range of the rate parameter λ.
- Normalize these likelihood values to create a probability distribution for λ.
- Incorporate uncertainty in the input data by repeating the process for the weighted alternative data sets and computing a weighted combination of the resulting probability distributions for λ.
- Represent the resulting probability distribution by a discrete approximation for implementation in hazard analysis.

The likelihood functions used to implement this process depend upon the type of paleoseismic data available. The recurrence data fall into two general categories. In the first, the data consist of an estimate of the number of earthquakes that have occurred after the deposit of a datable stratigraphic horizon, but estimates of the actual ages of the individual earthquakes are not available. These type of data are analogous to the assessments made for historical seismicity where the data consist of the number of earthquakes that have occurred within the period of catalog completeness. The second type of paleoseismic data are estimates of the actual ages of individual earthquakes. If these data are available, they provide estimates of the intervals between individual earthquakes that can be used to assess the average recurrence interval of

RLMEs. The following subsections describe the mathematical formulation used to assess earthquake recurrence rates from these two general types of paleoseismic data.

5.3.3.1.1 Earthquake Count in a Time Interval

The data available in this case is an estimate of the number of earthquakes, N, that have occurred in a known time interval T. For a Poisson model, the probability of observing N events in time interval T is given by

$$f(N) = \frac{(\lambda T)^N \exp\{-\lambda T\}}{N!} \tag{5.3.3-1}$$

This leads to the likelihood function for a rate parameter of λ_i given the observed data (N,T):

$$L(\lambda_i | N,T) = \frac{(\lambda_i T)^N \exp\{-\lambda_i T\}}{N!} \tag{5.3.3-2}$$

Figure 5.3.3-1 shows the relative likelihood based probability distribution developed using Equation 5.3.3-1 for the case where N equals two earthquakes and T equals 2,000 years. The resulting normalized uncertainty distribution for λ is in fact a gamma distribution with a mean value of (N+1)/T (e.g., Benjamin and Cornell, 1970). The mean differs from the maximum likelihood estimate of λ, which is N/T due to the skewness of the normalized likelihood distribution. In the case where the observed count is zero, the mean estimate of λ from the normalized likelihood function would be 1/T. This is analogous to the "natural floor" in the seismicity-based estimates of earthquake recurrence rates discussed in Section 5.3.2.1.1.

5.3.3.1.2 Earthquake Recurrence Intervals

The data available in this case are estimates of the occurrence times of a sequence of N+1 earthquakes obtained from dating of paleoseismic features or known occurrence times from the historical record. The estimated dates for the N+1 earthquakes provide N values of the times between earthquakes, t_i, plus the open interval since the most recent RLME, t_0. The Poisson process produces an exponential distribution for the time between occurrences:

$$f(t) = \lambda \exp\{-\lambda t\}$$
$$F(t) = 1 - \exp\{-\lambda t\} \tag{5.3.3-3}$$

This leads to the likelihood function for a rate parameter of λ_i given the observed data (t_i, i = 1, N, t_0):

$$L(\lambda_i | t_i, t_0) = \prod_{i=1}^{N} f(t_i) \times \left[1 - F(t_0)\right] = \lambda_i^N \exp\left\{-\lambda_i \times \left[\sum_{i=1}^{N} t_i + t_0\right]\right\} \tag{5.3.3-4}$$

The form of the likelihood function indicates that specification of the dates for the intermediate events is not needed as the term $\sum_{i=1}^{N} t_i + t_0$ is equal to the elapsed time since the oldest earthquake

in the sequence. The likelihood function given by Equation 5.3.3-4 is equivalent in form to that given by Equation 5.3.3-2 and leads to the same gamma-shaped uncertainty distribution for λ with mean $(N+1)\Big/\left(\sum_{i=1}^{N} t_i + t_0\right)$.

The maximum likelihood estimate of λ from Equation 5.3.3-4 is $N/\left(\sum_{i=1}^{N} t_i + t_0\right)$. This differs from a common practice of using the inverse of the average of the recurrence intervals due to the inclusion of the length of the open interval t_0 in the denominator. However, neglecting the open interval is ignoring information. The use of Equation 5.3.3-4 is considered more appropriate because it incorporates of all of the available information.

5.3.3.1.3 Representation of Continuous Distribution for λ by a Discrete Approximation

The logic tree structure is generally used to represent uncertainty in the inputs to a PSHA. In a logic tree, uncertainty distributions are represented by a (small) set of discrete alternatives with assigned probability weights. Miller and Rice (1983) provide a set of discrete approximations for arbitrary continuous probability distributions. Their five-point approximation was chosen to represent the distributions for λ developed for the RLME sources. The Miller and Rice (1983) five-point discrete approximation is given in Table 5.3.3-1. The first column indicates the location of the five discrete values for the variable in terms of their cumulative probability levels in the defined continuous distribution. These are illustrated by the five horizontal dashed lines in the lower panel of Figure 5.3.3-1. The second column lists the probability weights to be assigned to the five discrete values of the variable. The number of decimal points in the probability weights given in Miller and Rice (1983) perhaps implies too high a degree of precision. For application in the CEUS project, these probability weights are approximated by the values listed in the third column.

Application of the five-point approximation given above to the example shown on Figure 5.3.3-1 yields a discrete distribution with $\overline{\lambda} = 0.00149$ and $\sigma_\lambda = 0.000849$. The values for the continuous gamma distribution are $\overline{\lambda} = 0.0015$ and $\sigma_\lambda = 0.000866$, indicating that the discrete approximation does a good job of capturing the first two moments of the skewed continuous parameter distribution.

5.3.3.1.4 Use of Fault Slip Rates

In a few cases, the data used to estimate occurrence rates for RLMEs is in the form of fault slip rate. In these cases the value of λ is obtained by converting the fault slip rate into a moment rate and then dividing by the seismic moment of an RLME, i.e.:

$$\lambda = \frac{SA\mu}{\overline{M_O}(RLME)} \qquad (5.3.3\text{-}5)$$

where S is the fault slip rate, A is the total area of the fault surface, μ is crustal rigidity (taken to be 3×10^{11} dynes/cm^2), and $\overline{M_O}(RLME)$ is the average seismic moment of an RLME. The fault slip rate is computed by dividing the cumulative fault offset, D_{cum}, by the length of time over which it occurred, T:

$$S = \frac{D_{cum}}{T} \tag{5.3.3-6}$$

5.3.3.2 Estimation of Occurrence Rates for a Renewal Model

A renewal model for the occurrence probability of earthquakes is applied when the data suggest a more periodic behavior for the occurrence times between earthquakes than implied by the Poisson model. There are a variety of distributions that have been used to model the variability in the time between events, such as the lognormal, Weibull, and gamma distributions. Recently, Matthews et al. (2002) have proposed a model based on the inverse Gaussian distribution for inter-arrival times of repeated large ruptures on a fault. This model, termed the Brownian Passage Time (BPT) model was used by the Working Group (2003) to assess the probabilities of large earthquakes in the San Francisco Bay area. Ellsworth et al. (1999) and Matthews et al. (2002) propose that the BPT model is more representative of the physical process of strain buildup and release on a seismic source than the other distribution forms that have been used for renewal models (e.g., the lognormal). Based on these arguments, the BPT model was used in this analysis.

For the BPT model, the time interval between earthquakes, t, is distributed with probability density given by

$$f(t) = \left(\frac{\mu}{2\pi\alpha^2 t^3}\right)^{1/2} \exp\left(-\frac{(t-\mu)^2}{2\mu_{ln}\alpha^2 t}\right) \tag{5.3.3-7}$$

and cumulative probability given by

$$F(t) = \Phi[u_1(t)] + e^{2/\alpha^2}\Phi[-u_2(t)]$$

$$u_1(t) = \left(\sqrt{t/\mu} - \sqrt{\mu/t}\right)/\alpha \tag{5.3.3-8}$$

$$u_2(t) = \left(\sqrt{t/\mu} + \sqrt{\mu/t}\right)/\alpha$$

where μ is the mean inter-arrival time (repeat time), α is the aperiodicity coefficient (COV of t), and $\Phi(\ )$ is the standard normal cumulative probability function.

Given a sample of N time intervals and one open interval since the most recent earthquake, t_0, the likelihood function for the observed data set is given by

$$L(\mu,\alpha) = \prod_{i=1}^{N} f(t_i) \times \left[1 - F(t_0)\right] \tag{5.3.3-9}$$

with $f(t)$ and $F(t)$ given by Equation 5.3.3-7 and 5.3.3-8, respectively. Because of the very limited size of the data samples for the RLME sources where the model is applied, estimation of the aperiodicity coefficient, α, is highly uncertain. Therefore, the value of α is constrained to values reported from examination of larger data sets. Based on examination of a number of data sets, the Working Group (2003) developed an uncertainty distribution for the aperiodicity coefficient for the BTP model consisting of three weighted values for α of 0.3 (weight 0.2), 0.5 (weight 0.5), and 0.7 (weight 0.3). The Working Group (2003) weighted distribution was adopted to constrain the value of α.

The process described in Section 5.3.3.1 was used to develop a likelihood-based distribution for μ given a specified value of α. A broad range of values for μ were input into Equation 5.3.3-9 to compute likelihood values. These likelihoods were then normalized to form a probability distribution for μ for each of the three values of α. This distribution was then represented by the discrete five-point approximation from Miller and Rice (1983) described above.

For the renewal recurrence model, the probability of an earthquake in the next time interval Δt is given be the expression:

$$P_{renewal}(\text{event in time } t_0 \text{ to } t_0 + \Delta t) = \frac{F(t_0 + \Delta t) - F(t_0)}{1 - F(t_0)} \qquad (5.3.3\text{-}10)$$

The basic PSHA formulation used to assess the site hazard assumes that the occurrence of individual earthquakes conforms to a Poisson process. In order to combine the hazard from earthquakes defined by a renewal process into the total hazard, an equivalent Poisson rate is defined such that a Poisson process will give a probability of at least one earthquake in time interval Δt that is equal to the probability given by Equation 5.3.3-10. The equivalent Poisson rate, $\lambda_{renewal}$, is given by the expression:

$$\lambda_{renewal} = -\ln\left[1 - P_{renewal}(\text{event in time } t_0 \text{ to } t_0 + \Delta t)\right] / \Delta t \qquad (5.3.3\text{-}11)$$

A time period of 60 years was chosen as the time period of interest for application (20-year license life plus 40-year plant design life). The time t_0 was set at January 1, 2011. The corresponding values of $\lambda_{renewal}$ computed from the five weighted values of μ were then used as a discrete uncertainty distribution for the rate of RLMEs. Distributions are developed for each value of α, resulting in a total of 15 weighted alternative RLME rates.

5.3.3.3 Incorporating Uncertainty in the Input

The typical uncertainties encountered in characterizing the occurrence rates for RLME sources are uncertainties in the data set (i.e., specification of the number of earthquakes and the time span of their occurrence) and uncertainties in the actual ages of individual earthquakes or the age of the offset stratigraphic horizon used to define the time span.

Uncertainties in the appropriate data set are handled by creating a probability distribution for occurrence rate for each data set and then including these alternatives in the seismic source model logic tree as weighted alternatives. Examples of this are the alternative time periods of

complete paleoseismic reporting (e.g., 2,000 vs. 5,500 years) and the alternative number of RLME earthquakes that are evident in the data for the Charleston RLME, discussed in Section 6.1. A number of the other RLME source models also include alternative estimates of the number of RLMEs that have occurred in the recent past as discrete alternative models in the seismic source model logic tree.

Uncertainty in the age of individual earthquakes or the age of the offset marker horizon are typically expressed as a 90% or 95% confidence interval on the before present (BP) age, or as an age range (e.g., 10,000 to 12,000 years for the age of a marker horizon). This type of uncertainty is incorporated into the uncertainty distribution for the RLME rate parameter by simulating values of the age of the i^{th} RLME or of the total time span T from the expressed uncertainty distribution. The resulting values of t_i or T are then used to develop a likelihood based continuous probability distribution for the rate parameter. The process is then repeated multiple times and the resulting probability distributions averaged to produce a composite continuous distribution for the rate parameter. This composite distribution is then represented by the five-point discrete approximation described above.

For two of the RLME sources the available paleoseismic data consists of multiple samples providing constraints on the timing of past RLME occurrences. For the Charleston RLME, two types of data are used. One consists of samples whose ages are considered to be contemporary with the earthquake occurrence, denoted as CON ages. Each sample has its own age estimate with associated uncertainties. A distribution for the age of the associated RLME was developed by computing the relative likelihood that a range of possible ages produced the sample of age dates, assuming that the uncertainty in each age can be approximated by a normal distribution with the specified uncertainty ranges. These relative likelihoods were then normalized to produce a probability distribution for the RLME age. Figure 5.3.3-2 shows the resulting distributions for the ages of the five most recent Charleston RLMEs preceding the 1886 earthquake, labeled as "Con Ages."

The second type of sample data for the Charleston RLME is those that provide constraints on the event age in terms of representing points in time either before or after each earthquake occurrence. This additional age data was used to modify the CON age distributions by simulating ages from the CON age distribution and from the distribution for the constraining sample age and rejecting those simulations that violate the age constrains. The resulting earthquake age distributions are shown on Figure 5.3.3-2 as the "All Ages" distributions.

These two sets of age distributions were then used to develop uncertainty distributions for the RLME occurrence rate.

5.3.3.4 RLME Magnitude Distribution

The RLME sources are intended to model the repeated occurrence of large earthquakes of similar size. The expected magnitude of the RLME is estimated from various sources of data and its uncertainty is expressed by a probability distribution in the RLME source logic tree. In addition, it is anticipated that there will be variability in the actual magnitude of each earthquake when it occurs, requiring the need for specifying an aleatory uncertainty for earthquake magnitude. For

this study, a simple uniform distribution is assumed, which implies maximum uncertainty. The width of this distribution is set to be the expected RLME magnitude ± ¼-magnitude units, resulting in a range of ½-magnitude units. A ½-magnitude unit range implies a factor of 3 range in source dimensions and over a factor of 5 range in seismic moment. The ½-magnitude range was chosen as a typical representation of the variability of repeated large earthquakes occurring on a single source (e.g., Youngs and Coppersmith, 1985).

5.4 Assessment of Future Earthquake Characteristics

The CEUS SSC model has been developed for use in future PSHAs. To make this future use possible, the SSC model must be combined with a ground-motion characterization (GMC) model. At present, the ground-motion prediction equations (GMPEs) in use for SCRs such as the CEUS include limited information regarding the characteristics of future earthquakes. In contrast, GMPEs for active tectonic environments often include multiple characteristics, including style of faulting as defined by sense of fault slip and distance to the fault rupture. In anticipation of the possible future development of GMPEs for the CEUS that will make it possible to incorporate similar types of information, a number of characteristics of future earthquakes in the CEUS are assessed in this section. In addition to characteristics that might be important for ground-motion assessments, there are also assessed characteristics that are potentially important to the modeling conducted for hazard analysis. For example, assuming that the PSHA models earthquakes as finite fault ruptures, information is then needed about the relationship between magnitude of the earthquake and rupture dimensions, length-to-width aspect ratios, and the relationship between the rupture and the seismic source zone boundaries.

The assessment by the TI Team of future earthquake characteristics was carried out in two parts. In the first, a set of characteristics believed to be representative of the entire CEUS study region was assessed. This set of characteristics is termed “default” because, in the absence of local data for a particular region, it would provide a basis for use in the PSHA calculations. The default characteristics and their assessed values are given in Table 5.4-1. The second assessment involved making estimates of the future earthquake characteristics for the individual seismic sources. These assessments were made by the TI Team by evaluating source-specific data. A summary of the source-specific assessments is given in Table 5.4-2. A brief description of the future earthquake characteristics is given below.

The assessment of future earthquake characteristics for the entire CEUS study region is given in Table 5.4-1. Also included in Table 5.4-1 are representative references to data and interpretations that have been evaluated in the assessment of these characteristics. It should be noted that these references usually cite local or site-specific studies, and not evaluations across the entire CEUS SSC study region, because such studies are not available on a regional scale. In addition, the references only provide data for consideration by TI Team members in their assessments of the future earthquake characteristics; the assessed values in column 2 of the table are based on the assessments by the TI Team of the default characteristics that represent the current state of scientific knowledge” for the CEUS in the absence of source-specific information.

In the characterization of individual seismic sources, if source-specific data suggest a difference from the default case, the characteristics were assessed for that particular source. The source-

specific assessments summarized in Table 5.4-2 are discussed for each RLME source and Mmax source in Chapter 6 and for each seismotectonic zone in Chapter 7. The discussions in Chapters 6 and 7 include a description of the assessments of the future earthquake characteristics, as well as the technical bases for the weights applied to the alternatives. As indicated in Table 5.4-2, those assessments that are the same as the default assessments are shown as italics; those assessments that are different from the default assessments due to source-specific information are shown in plain text.

Subsequent to the finalization of the model for developing the HID and hazard calculations, minor revisions to the future earthquake characteristics were made to the ECC-AM and ECC-GC sources. The final characteristics are given in Table 5.4-2. These characteristics have little to no effect on the hazard, given the ground motion models that were used for the hazard calculations. However, future users of the CEUS SSC model should consider the characteristics given in Table 5.4-2 to be the final assessments for the CEUS SSC model.

The Data Evaluation table for the assessment of default characteristics of future earthquakes in the CEUS is presented in Appendix Table C-5.4, and the Data Summary table in Appendix Table D-5.4. Each of the future earthquake characteristics is described below.

5.4.1 Tectonic Stress Regime

Some GMPEs provide different relationships as a function of compressional or extensional tectonic stress regimes. A variety of data sets, including the stress data compiled for the project by Zoback (2010), indicate the CEUS study region lies within a contemporary tectonic stress regime that is compressional.

5.4.2 Sense of Slip/Style of Faulting

Sense of slip of future earthquakes, which defines the style of faulting, is assessed based on a variety of data sets, including earthquake focal mechanisms, stress indicators, and the relationship between stress orientations and the geologic structure. The result of the assessment given in Table 5.4-1 is aleatory variability, which reflects the expected relative frequency of strike-slip versus reverse faulting averaged across the study region.

5.4.3 Strike and Dip of Ruptures

It is assumed that the GMPE for the CEUS will be capable of incorporating information about the characteristics of earthquake sources, modeled as finite faults in much the same manner as earthquake sources ares modeled in the WUS. Therefore, Table 5.4-1 includes a number of assessment results that address the expected geometry of future ruptures and, in some cases, the relationship of the geometry to earthquake magnitude. A key attribute of the geometry is the strike and dip (i.e., orientation) of the rupture surfaces. Assessment of the strike of future rupture surfaces was made using information from earthquake focal mechanisms, the structural grain of major fold and fault systems—unless the grain is judged to have no relationship to contemporary earthquake properties; and tectonic stress data in regard to the orientations and magnitudes of

principal stresses. Assessment of the dip of future rupture surfaces was dependent on style of faulting and derived from empirical data related to observed ruptures and evaluations of physical properties.

5.4.4 Seismogenic Crustal Thickness

From the standpoint of high-frequency energy release and seismic hazard, an important consideration is the geometry of the rupture within the seismogenic part of the crust. The thickness of the seismogenic crust (also called the seismogenic layer or seismogenic zone) varies spatially, but the results of the assessment shown in Table 5.4-1 represent an epistemic assessment of the average crustal thickness across the study region. This distribution can be modified locally to account for unusually thick or thin seismogenic crust at the scale of an individual seismic source. The assessment of seismogenic crustal thickness is based on the focal depth distributions of well-studied earthquakes, usually those lying within local or regional seismic networks, such that hypocentral depth can be determined with a high degree of confidence. However, uncertainties in focal depth across the study region vary considerably as a function of available hypocenter information.

Given a focal depth distribution, physical considerations given by various researchers suggest that the base of the seismogenic zone is identified as lying near the base of the observed focal depths. For example, Scholz (1998) identifies the 300°C isotherm as corresponding to the onset of dislocation creep in quartz, which he interprets to control the seismic/aseismic transition zone by Scholz (1998). Tanaka (2004) and Tanaka and Ito (2002) compares high-quality thermal measurements and seismicity depth data to examine the conept that temperature is a fundamental parameter for determining the thickness of the seismogenic zone. Their gridded heat flow or geothermal gradient and D_{90}, the depth above which 90% of earthquakes occur, correlated well with each other. The evaluated temperatures for D_{90} range between 250°C and 450°C. The consistency of temperature for D_{90} over a large depth interval almost all over the Japanese Islands leads these researchers to conclude that temperature is the dominant factor governing the focal depth in the crust. However, other evaluators have used criteria, such as the 95th percentile of the focal depth distribution. Given these alternative interpretations, and the uncertainty in focal depth assessments across the the study region, the generic and zone-specific assessments of crustal thickness presented in Tables 5.4-1 and 5.4-2 were made.

As discussed in Chapter 8, uncertainty in crutal thickness is not a significant contributor to uncertainty in seismic hazard in the CEUS because the current ground motion prediction equations have only limited sensitivity to depth. However, in anticipation of future advances in the characterization of ground motions for CEUS earthquakes, additional analysis of focal depth data are presented. Given the assumption that the criterion of D_{90} is the correct interpretation of the base of the seismogenic layer, a high-quality focal depth data set was compiled based on those earthquakes having the best-resolved focal depths. The calculation of D_{90} is made for each of the seismic source zones. The assessment of seismogenic thickness and its uncertainty is, in turn, a function of the number of earthquake focal depths that exist for each source. A bootstrapping approach was used that uses the assumption that the random uncertainty in the focal depth of any given earthquake is about ±2 km and the uncertainty in the actual thickness is a function of the number of number of earthquakes that define D_{90}.

Figure 5.4.4-1 shows the spatial distribution of earthquakes in the CEUS SSC Project catalog with $E[\mathbf{M}]$ magnitudes of 2.2 and greater. Figure 5.4.4-2 shows the spatial distribution of earthquakes in the CEUS SSC Project database having the best-resolved focal depths. This include smaller earthquakes for which $E[\mathbf{M}]$ magnitudes were not computed. In general, the spatial distribution of earthquakes with better quality depth determinations is similar to that of the full catalog. However, some areas, such as the Gulf Coast region and the upper Great Plains have limited coverage.

Figure 5.5.4-3 shows histograms of focal depth for earthquakes within the Mmax source zones and Figure 5.5.4-4 shows histograms of focal depth for earthquakes within the seismotectonic source zones. Several of the seimotectonic source zones have limited to very limited data (AHEX, ECC-GC, GHEX, and OKA). Assessment of D_{90} for these sources was therefore based on assumed similarity with other sources. As discussed in Section 3.2.4, an attempt was made to identify fixed depth earthquakes. Dips in the histograms for a number of source zones at depths of 5, 10, and 18 km indicates that for some of the earthquakes flagged fixed depths may not be fixed. However, adding synthetic data to representing estimates of the number of missing earthquakes did not result in a significant difference in the estimate of D_{90}.

Bootstrap samples of focal depths were drawn from the observed focal depth distributions for each source zone. A normally distributed random depth error with a standard deviation of 2 km was added to each focal depth. The bootstrap samples were then used to compute an uncertainty distribution for D_{90}. Table 5.4-3 lists the results of the analyses. The uncertainty range for D_{90} is typically ±1 km about the observed value. Three of the source zones (AHEX, GHEX, and OKA have too few good quality focal depths to make an assessment and the results for ECC-GC may also be too broad due to the limited sample size.

The results presented in Table 5.4-3 can be used as part of use of the CEUS SSC model in site-specific applications and to incorporate potential advancements on ground motion modeling.

5.4.5 Fault Rupture Area

In the hazard analysis, each event is associated with a moment magnitude from the recurrence relationship and, as a result, can be associated with a rupture area that is a function of that magnitude. A variety of empirical relationships between magnitude and rupture area have been proposed in the literature, most leading to very similar results for crustal earthquakes. Because of those similarities, it was judged sufficient to use a single relationship to represent the range of estimates. The relationship used (Somerville et al., 2001) was developed specifically for application in eastern North America, thus it is deemed to be applicable to the CEUS for this assessment. In the future, consideration should be given for updating this assessment with more recent relationships such as that given by Leonard (2010).

5.4.6 Rupture Length-to-Width Aspect Ratio

Given a rupture area, the shape of the rupture must be defined within the seismogenic crust. Based on an evaluation of observed ruptures for moderate-to-large earthquakes, an assessment

was made of the length-to-width aspect ratio. For smaller rupture areas, the aspect ratio is assessed to be equi-dimensional or one-to-one. For progressively larger rupture areas, the downdip width of the rupture surface, which varies depending on dip angle, will eventually involve the entire seismogenic crustal thickness. At still larger rupture areas, only rupture length will increase, leading to progressively larger aspect ratios. Such a model will result in large aspect ratios for large strike-slip earthquakes and somewhat lower aspect ratios for the same-magnitude reverse earthquakes, which is consistent with empirical observations. For the smaller equi-dimensional ruptures, the depth of the rupture surface is related to the focal depth distribution, following the approach given in NAGRA (2004, p. 174). That approach uses the focal depth distribution as a basis for the relative likelihood of the center pointof ruptures occurring at any particular depth. Given the small magnitudes of these earthquakes, the effect of this assumption on hazard is small.

5.4.7 Relationship of Rupture to Source Zone Boundaries

For a finite-fault SSC model, the locations of future earthquakes are defined by an epicenter and an associated rupture, the geometry of which is represented by the assessment results as discussed above in Sections 5.4.3 through 5.4.6. The event epicenters lie within the seismic source zone of interest, but it is also possible that the associated ruptures may propagate outside the source zone, depending on the properties assessed for the source boundary. If the source boundary is assessed as “leaky,” then ruptures will be allowed to extend beyond the boundary; if they are “strict,” the ruptures will be confined within the boundary. The earthquake epicenter is assumed to lie at the midpoint of the rupture length. Therefore, a leaky rupture would entail no more than 50% of the rupture length (for the particular case where the epicenter is located at the zone boundary). The assessment of whether a boundary is leaky or strict is related to the orientation of rupture surfaces and the expected orientation of ruptures in the adjacent seismic source zones. Sharp source boundaries defined by distinct changes in expected rupture geometries are likely to be strict boundaries. For example, a boundary of a seismic source zone that marks the distinct change in orientation of future earthquake ruptures would likely be assessed as having a strict boundary so that ruptures from one source with a particular orientation are not modeled with that orientation in the adjacent source. In all cases, the lengths of ruptures that are assessed to occur within seismic sources are checked to be sure that their rupture dimensions would not exceed the dimensions allowed within the source (and rupture across the source boundary, as applicable) for the earthquake magnitudes that are assessed. Across the region, the general continuity of expected rupture orientations results in the default assessment of leaky boundaries for most seismic source zones.

5.5 Predicted Seismic Moment Rate

The CEUS SSC seismic source model developed in Chapters 4, 5, and 6 provides an assessment of earthquake recurrence rates for the entire CEUS study region. These recurrence rates were used to compute the predicted mean seismic moment rate for the study region. The table below summarizes these results. Two thirds of the seismic moment rate results from the recurrence of New Madrid RLMEs. The remaining contribution is evenly split between the remaining RLME sources and the distributed seismicity sources.

Predicted Seismic Moment Rate for CEUS SSC Model

Source	Seismic Moment Rate (dyne-cm/year)
Distributed Seismicity	3.2×10^{24}
New Madrid RLME	1.0×10^{25}
Remaining RLMEs	3.4×10^{24}
Total	1.7×10^{25}

Johnston et al. (1994) assessed a seismic moment rate for the SCR region of North America both including and excluding the New Madrid 1811-1812 earthquake sequence. Their assessment is listed in the table below in terms of seismic moment rate per 10^5 km^2. Also listed are the mean moment rates predicted using the CEUS SSC model and the area encompassed by the CEUS SSC seismic source model of 7.8×10^6 km^2. The results are similar excluding the New Madrid earthquakes and with a factor of 2 including the New Madrid earthquakes. Johnston et al. (1994) used moment magnitudes **M** 8.1–8.3. The characterization of the New Madrid RLME for the CEUS SSC model results in a mean magnitude for these earthquakes in updated magnitude assessments for these earthquakes s for these earthquakes are approximately **M** 7.5. The one-half magnitude difference in the moment magnitudes assigned to these earthquakes translates into approximatelty a factor of 5 in seismic moment. Thus, the difference in seismic moment rate assigned to the New Madrid sequence based on magnitude assessments for the three earthquakes easily accounts for differences in the estimated moment rate for the CEUS. Based on these comparisons it is concluded that the seismic moment rate predicted by the CEUS SSC model is consistent with the values estimated by Johnston et al. (2004).

Seismic Moment Rates for the SCR Portion of North America

Comparison Case	Seismic Moment Rate Density (dyne-cm/year/10^5 km^2)	
	Johnston et al. (1994)	CEUS SSC Model
Including New Madrid	4.5×10^{23}	2.2×10^{23}
Excluding New Madrid	1.07×10^{23}	0.8×10^{23}

Table 5.2.1-1
Mesozoic-and-Younger Extended Superdomains (MESE)

Superdomain Number	Total Area (km^2)	No. of Domains	Type	Age	Stress	Source Stress Angle	$m_{\max-obs}$ From Catalog	Number of Earthquakes M ≥ 4.5
SE01*	1326660	6	EC	CZ	C	F	6.3	65.8
SE02*	4369994	15	EC	CZ	C	NF	7.9	871.7
SE03*	2125357	7	EC	CZ	C	UK	7.3	32.9
SE04	1651160	1	EC	CZ	E	NF	6.2	16.1
SE05	2149008	4	EC	CZ	U	UK	5.2	4
SE06*	4057142	10	EC	MZ	C	F	7.3	63.2
SE07*	5323365	21	EC	MZ	C	NF	7.8	557.2
SE08*	3638444	8	EC	MZ	C	UK	6.5	21.9
SE09	1272700	1	EC	MZ	E	F	6.2	4
SE10*	1279577	4	EC	MZ	E	NF	5.9	13
SE11	1196907	2	EC	MZ	E	UK	5.2	1
SE12	4314292	14	EC	MZ	U	UK	5.6	8
SE13	31132	1	EC	PC	C	F	0	0
SE14	507439	2	EC	PC	C	NF	5.5	8.5
SE15	22629	1	EC	PC	E	UK	6.5	2
SE16	556326	3	EC	PC	U	UK	5.4	1
SE17	82782	1	EC	PZ	C	F	5.5	4.1
SE18	570732	4	EC	PZ	C	NF	6.5	8.2
SE19	27315	1	EC	PZ	C	UK	5.3	3
SE20	1257982	3	EC	PZ	U	UK	6.1	11.3

* Indicates superdomain used to develop prior distribution.

C—Compressive
CZ—Cenozoic
E—Extensive
EC—Extended crust
F—Favorable orientation
MZ—Mesozoic
NE—Non-extended crust
NF—Unfavorable orientation
PC—Precambrian
PZ—Paleozoic
U—Unknown

Table 5.2.1-2
Older Extended and Non-Extended Superdomains (NMESE)

Superdomain Number	Total Area (km^2)	No. of Domains	Type	Age	Stress	Source Stress Angle	$m_{\max-obs}$ From Catalog	Number of Earthquakes M ≥ 4.5
SE13	31132	1	E	PC	C	F	0	0
SE14*	507439	2	E	PC	C	NF	5.5	8.5
SE15	22629	1	E	PC	E	UK	6.5	2
SE16	556326	3	E	PC	U	UK	5.4	1
SE17	82782	1	E	PZ	C	F	5.5	4.1
SE18*	570732	4	E	PZ	C	NF	6.5	8.2
SE19	27315	1	E	PZ	C	UK	5.3	3
SE20	1257982	3	E	PZ	U	UK	6.1	11.3
SN01	33900	1	NE	MZ	C	F	0	0
SN02*	805784	2	NE	MZ	C	NF	6.2	46.5
SN03*	835876	2	NE	MZ	C	UK	6.4	92.6
SN04	599472	3	NE	MZ	U	UK	5	2
SN05*	9759298	16	NE	PC	C	F	6.7	225.2
SN06*	19633024	17	NE	PC	C	NF	7.3	277.8
SN07*	34329412	38	NE	PC	C	UK	6.8	765.6
SN08*	1519341	3	NE	PC	E	F	5.7	22
SN09*	1689145	3	NE	PC	E	NF	6.6	401.2
SN10*	2039146	3	NE	PC	E	UK	5.6	13.4
SN11	9474115	20	NE	PC	U	UK	6.4	37.1
SN12*	2592956	7	NE	PZ	C	F	6.4	294.4
SN13*	1870538	6	NE	PZ	C	NF	6.4	75.9
SN14*	6518060	16	NE	PZ	C	UK	6.8	110.4
SN15	3645521	9	NE	PZ	U	UK	5.7	12.7

* Indicates superdomain used to develop prior distribution.

C—Compressive
CZ—Cenozoic
E—Extensive
EC—Extended crust
F—Favorable orientation
MZ—Mesozoic
NE—Non-extended crust
NF—Unfavorable orientation
PC—Precambrian
PZ—Paleozoic
U—Unknown

Table 5.2.1-3
Composite SCR Superdomains (COMP)

Superdomain Number	Total Area (km^2)	Number of Domains	Type	Age	Stress	Source Stress Angle	$m_{\max-obs}$ From Catalog	Number of Earthquakes M ≥ 4.5
SD01*	1326660	6	NA	CZ	C	F	6.3	65.8
SD02*	4369994	15	NA	CZ	C	NF	7.9	871.7
SD03*	2125357	7	NA	CZ	C	UK	7.3	32.9
SD04	1651160	1	NA	CZ	E	NF	6.2	16.1
SD05	2149008	4	NA	CZ	U	UK	5.2	4
SD06*	4091042	11	NA	MZ	C	F	7.3	63.2
SD07*	6129149	23	NA	MZ	C	NF	7.8	603.7
SD08*	4474320	10	NA	MZ	C	UK	6.5	114.5
SD09	1272700	1	NA	MZ	E	F	6.2	4
SD10*	1279577	4	NA	MZ	E	NF	5.9	13
SD11	1196907	2	NA	MZ	E	UK	5.2	1
SD12	4913764	17	NA	MZ	U	UK	5.6	10
SD13*	9790430	17	NA	PC	C	F	6.7	225.2
SD14*	20140460	19	NA	PC	C	NF	7.3	286.3
SD15*	34329412	38	NA	PC	C	UK	6.8	765.6
SD16*	1519341	3	NA	PC	E	F	5.7	22
SD17*	1689145	3	NA	PC	E	NF	6.6	401.2
SD18*	2061775	4	NA	PC	E	UK	6.5	15.4
SD19	10030441	23	NA	PC	U	UK	6.4	38.1
SD20*	2675738	8	NA	PZ	C	F	6.4	298.5
SD21*	2441270	10	NA	PZ	C	NF	6.5	84.1
SD22*	6545375	17	NA	PZ	C	UK	6.8	113.4
SD23	4903503	12	NA	PZ	U	UK	6.1	24

* Indicates superdomain used to develop prior distribution.

C—Compressive
CZ—Cenozoic
E—Extensive
EC—Extended crust
F—Favorable orientation
MZ—Mesozoic
NE—Non-extended crust
NF—Unfavorable orientation
PC—Precambrian
PZ—Paleozoic
U—Unknown

Table 5.2.1-4
Results of Analyses of Updated SCR Superdomains

Superdomain Set	Average Number of Earthquakes ≥4.5	Average *b*-value	Mean $m_{\max-obs}$	Standard Deviation	Bias Adjusted Mean m^u
MESE	232	0.85	7.05	0.75	7.35
NMESE	180	1.02	6.48	0.61	6.70
COMP	248	0.94	6.88	0.64	7.20

Table 5.2.1-5
Source Zones, P(m^u > 8¼) Values, and Weights on Kijko (2004) K-S-B Estimates

Source	P(m^u > 8¼)	Weight Assigned to Kijko Approach
Study Region (paleo eq)	0.84	0
MESE-N	0.44	0.06
MESE-W (paleo eq)	0.83	0
NMESE-N (paleo eq)	0.87	0
NMESE-W	0.27	0.23
AHEX	1	0
ECC-AM	0.45	0.05
ECC-GC	0.78	0
GHEX	0.8	0
GMH	0.81	0
IBEB (paleo eq)	0.72	0
MidC-A	0.26	0.24
MidC-B	0.26	0.24
MidC-C	0.26	0.24
MidC-D	0.26	0.24
NAP	0.58	0
OKA	0.69	0
PEZ-N	0.65	0
PEZ-W	0.49	0.01
RR	0.67	0
RR-RCG	0.43	0.07
SLR	0.48	0.02

Note: "Paleo eq" indicates that the zone uses paleoseismic evidence to constrain the size of the largest observed earthquake in that zone. In those cases, the Kijko approach is assigned zero weight.

Table 5.2.1-6
Mmax Distributions for the Two Example Seismic Sources

Weight	Maximum Magnitude for:	
	MESE-N	**NAP**
0.101	6.1	5.8
0.244	6.3	6.4
0.310	6.6	6.9
0.244	7.0	7.5
0.101	7.8	8.0

Table 5.3.2-1
Alternative Cases Considered for the Magnitude-Dependent Weights

Case	M 2.9–3.6	M 3.6–4.3	M 4.3–5.0	M 5.0–5.7	M 5.7–6.4	M > 6.4
A (wt = 0.3)	1	1	1	1	1	1
B (wt = 0.3)	0.1	1	1	1	1	1
C	0	1	1	1	1	1
D	0	0.1	1	1	1	1
E (wt = 0.4)	0	0.3	1	1	1	1

Note: The cases used in the final calculations and the weights assigned to them are shown in bold.

Table 5.3.3-1
Miller and Rice (1983) Discrete 5-Point Approximation to a Continuous Probability Distribution and the Modified Form Used in This Study

Cumulative Probability Level of Continuous Distribution	Assigned Probability Weight	Modified Probability Weight
0.034893	0.10108	0.101
0.211702	0.24429	0.244
0.5	0.30926	0.310
0.788298	0.24429	0.244
0.965107	0.10108	0.101

Table 5.4-1
Assessment of Default Characteristics of Future Earthquakes in the CEUS

Future Earthquake Characteristic	TI Team Assessments (Relative frequencies or weighted alternatives)	References[1]
Tectonic Stress Regime	Compressional	Zoback (2010)
Sense of Slip/Style of Faulting	Treat as aleatory (relative frequency): • 2:1 strike-slip:reverse	Focal mechanisms: New Madrid (Shumway, 2008; Horton et al., 2005) Eastern Tennessee (Chapman et al., 1997) Central Virginia (Kim and Chapman, 2005) St. Lawrence (S. Mazzotti, Wksp 2) Southern Great Lakes (Dineva et al., 2004) Wabash Valley (Kim, 2003) Pennsylvania (Seeber et al., 1998) CEUS (Zoback, 2010; Heidbach et al., 2008; van Lanen and Mooney, 2007) Charleston (P. Talwani, Wksp 2) Zoback (1992)
Strike and Dip of Ruptures	Aleatory distribution: • N50W (0.2) • N-S (0.2) • N35E (0.4) • N60E (0.1) • E-W (0.1) Dip is a function of sense of slip: • Strike-slip (90°–60°) (uniform) • Reverse (30°–60°) (uniform) • Either direction (50:50)	van Lanen and Mooney (2007) Sibson and Xie (1998) Zoback (1992) Marshak and Paulsen (1997) (NW trends)

Future Earthquake Characteristic	TI Team Assessments (Relative frequencies or weighted alternatives)	References[1]
Seismogenic Crustal Thickness	Epistemic distribution: • 13 km (0.4) • 17 km (0.4) • 22 km (0.2)	Sibson (1984, 2007) van Lanen and Mooney (2007) for ENA Mai et al. (2005) Atkinson (2004a) Sykes et al. (2008) P. Talwani (Wksp 2)
Fault Rupture Area	Function of magnitude: • Use Somerville et al. relation for ENA	Somerville et al. (2001)
Rupture Length-to-Width Aspect Ratio	Function of rupture area: • 1:1 for smaller ruptures • With progressively larger areas, when rupture width equals seismogenic crustal thickness, extend only the length	Wesnousky (2008) Toro approach in NAGRA (2004)
Relationship of Rupture to Source Zone Boundaries	• Epicenter is at center of rupture length (map view) • All boundaries are “leaky”; rupture is allowed to extend beyond boundary. (Note: If boundary is “strict,” rupture cannot extend beyond boundary, although epicenter can be near boundary)	

1. References provide insight into the assessment of characteristics; however, they do not uniquely define them for purposes of the CEUS SSC study. Those assessments were made by the TI Team.

Table 5.4-2
Characteristics of Future Earthquakes for Individual Seismic Sources

Source	Sense of Slip[1]	Rupture Strike[1]	Rupture Dip[1]	Source Boundaries	Seismogenic Crustal Thickness[2]
RLME Sources					
Charlevoix	Reverse	Uniform 0°–360°	Uniform 40°–60°	Leaky	25 km (0.8) 30 km (0.2)
Charleston —Large	Strike-slip	NE parallel to long axis (0.8) NW parallel to short axis (0.2)	90°	Strict	*13 km (0.4)* *17 km (0.4)* *22 km (0.2)*
Charleston —Local	Strike-slip	NE parallel to long axis	90°	Strict	*13 km (0.4)* *17 km (0.4)* *22 km (0.2)*
Charleston —Narrow	Strike-slip	NE parallel to long axis	90°	Leaky at ends	*13 km (0.4)* *17 km (0.4)* *22 km (0.2)*
Cheraw	Normal-oblique	On fault trace (NE)	50° NW (0.6) 65° NW (0.4)	Strict	*13 km (0.4)* *17 km (0.4)* *22 km (0.2)*
Commerce	Strike-slip	NE parallel to long axis of zone	90°	Leaky at ends	13 km (0.4) 15 km (0.4) 17 km (0.2)
ERM-N	Strike-slip	NE parallel to long axis of zone	90°	Leaky at ends	13 km (0.4) 15 km (0.4) 17 km (0.2)
ERM-S	Strike-slip	NE parallel to long axis of zone	90°	Leaky at ends	13 km (0.4) 15 km (0.4) 17 km (0.2)
Marianna	Strike-slip	NE 45° (0.5) NE 45° (0.5)	90°	Leaky at ends	13 km (0.4) 15 km (0.4) 17 km (0.2)
Meers—Fault	Strike-slip (0.5) Reverse (0.5)	On fault	Strike-slip 90° Reverse 40° SW	Strict	15 km (0.5) 20 km (0.5)
Meers—Random in OKA	Reverse oblique	Parallel to long axis of zone	Uniform 40°–90°	Strict	15 km (0.5) 20 km (0.5)
NMFS	NMN, NMS: Strike-slip RMT: reverse	On fault	NMN, NMS: 90° RFT: 40° SW	Strict	13 km (0.4) 15 km (0.4) 17 km (0.2)

Source	Sense of Slip[1]	Rupture Strike[1]	Rupture Dip[1]	Source Boundaries	Seismogenic Crustal Thickness[2]
Wabash Valley	*2/3 Strike-slip* *1/3 Reverse*	Strike parallel to the long axis of the zone (0.8) N50W (0.1) N20W (0.1)	2/3 Strike-slip, 90° 1/3 Reverse, 40°–60° Strike-slip, 90° Reverse, 40°	*Leaky*	17 km (0.7) 22 km (0.3)
Seismotectonic Zones					
AHEX	*2/3 Strike-slip* *1/3 Reverse*	*N50W (0.1)* *N-S (0.1)* *N25E (0.4)* *N60E (0.3)* *E-W (0.1)*	*Strike-slip (90°–60°) (uniform)* *Reverse (30°–60°) (uniform)*	*Leaky*	8 km (0.5) 15 km (0.5)
ECC-AM	*2/3 Strike-slip* *1/3 Reverse*	*N50W (0.2)* *N-S (0.2)* *N35E (0.4)* *N60E (0.1)* *E-W (0.1)*	*Strike-slip (90°–60°) (uniform)* *Reverse (30°–60°) (uniform)*	*Leaky*	*13 km (0.6)* *17 km (0.3)* *22 km (0.1)*
ECC-GC	*2/3 Strike-slip* *1/3 Reverse*	Uniform 0° to 360°	*Strike-slip (90°–60°) (uniform)* *Reverse (30°–60°) (uniform)*	*Leaky*	13 km (0.6) 17 km (0.3) 22 km (0.1)
GHEX	*2/3 Strike-slip* *1/3 Reverse*	Uniform 0° to 360°	*Strike-slip (90°–60°) (uniform)* *Reverse (30°–60°) (uniform)*	*Leaky*	8 km (0.5) 15 km (0.5)
GMH	*4/5 Reverse* *1/5 Strike-slip*	N50W (0.4) N20W (0.4) E-W (0.2)	*Strike-slip (90°–60°) (uniform)* *Reverse (30°–60°) (uniform)*	Leaky	25 km (0.5) 30 km (0.5)
IBEB	SS (0.1) RO (0.1) R (0.3) SS (0.2) SS (0.3)	N50W (0.1) N20W (0.1) N-S (0.3) E-W (0.2) N40E (0.3)	90° 70 E (0.5), 70 W (0.5) 40E/W (0.4);70E/W (0.3) 90° 90°	Strict	*13 km (0.4)* *17 km (0.4)* *22 km (0.2)*

Source	Sense of Slip[1]	Rupture Strike[1]	Rupture Dip[1]	Source Boundaries	Seismogenic Crustal Thickness[2]
PEZ	*2/3 Strike-slip* *1/3 Reverse*	N50W (0.2) N-S (0.2) N35E (0.4) N60E (0.1) E-W (0.1)	*Strike-slip (90°–60°) (uniform* *Reverse (30°–60°) (uniform)*	Leaky	*13 km (0.4)* *17 km (0.4)* *22 km (0.2)*
MidC	*2/3 Strike-slip* *1/3 Reverse*	*N50W (0.2)* *N-S (0.2)* *N35E (0.4)* *N60E (0.1)* *E-W (0.1)*	*Strike-slip (90°–60°) (uniform)* *Reverse (30°–60°) (uniform)*	Strict	*13 km (0.4)* *17 km (0.4)* *22 km (0.2)*
NAP	*1/3 Strike-slip* *2/3 Reverse*	N50W (0.2) N-S (0.2) N35E (0.4) N60E (0.1) E-W (0.1)	*Strike-slip (90°–60°) (uniform)* *Reverse (30°–60°) (uniform)*	Leaky	*13 km (0.4)* *17 km (0.4)* *22 km (0.2)*
OKA	Reverse Oblique	Parallel to long axis of zone	Uniform 45°–75°	*Leaky*	15 km (0.5) 20 km (0.5)
RR and RR-RCG	SS (0.2) R (0.35) SS (0.5) SS (0.2) SS (0.2)	N50W (0.2) N10W(0.35) E-W (0.05) N30E (0.2) N55E (0.2)	90° 70 E/W (0.5), 40 E/W (0.5) 90° 90° 90°	Strict	13 km (0.4) 15 km (0.4) 17 km (0.2)
SLR	*1/3 Strike-slip* *2/3 Reverse*	N25E (0.2) N40E (0.2) N70E (0.2) N50W (0.15) N70W (0.15) N-S (0.05) E-W (0.05)	*Strike-slip (90°–60°) (uniform* *Reverse (30°–60°) (uniform)*	Leaky	25 km (0.5) 30 km (0.5)

Note: Default characteristics are indicated in italics.

1. Weights reflect aleatory variability; weights are therefore relative frequencies.

2. Weights reflect epistemic uncertainty; weights are therefore relative credibility that the given thickness is correct.

Table 5.4-3
Estimates of D_{90} for Individual Seismic Source Zones

Seismic Source Zone	5th Percentile D_{90} (km)	D_{90} from Observed Focal Depths (km)	95th Percentile D_{90} (km)
MESE-N	16	17	18
MESE-W	17	18	19
NMESE-N	17	18	19
NMESE-W	15	16	17
AHEX	—	—	—
ECC-AM	11	12	13
ECC-GC	12	16	20
GHEX	—	—	—
GMH	22	23	24
IBEB	18	19	21
MIDC-A	15	16	17
MIDC-B	15	16	17
MIDC-C	15	16	17
MIDC-D	15	16	17
NAP	15	17	20
OKA	—	—	—
PEZ-N	18	19	20
PEZ-W	18	19	20
RR	12	13	14
RR-RCG	12	13	14
SLR	19	20	21

Note: Cells with "—" indicate source zones with too few earthquakes from which to obtain an estimate.

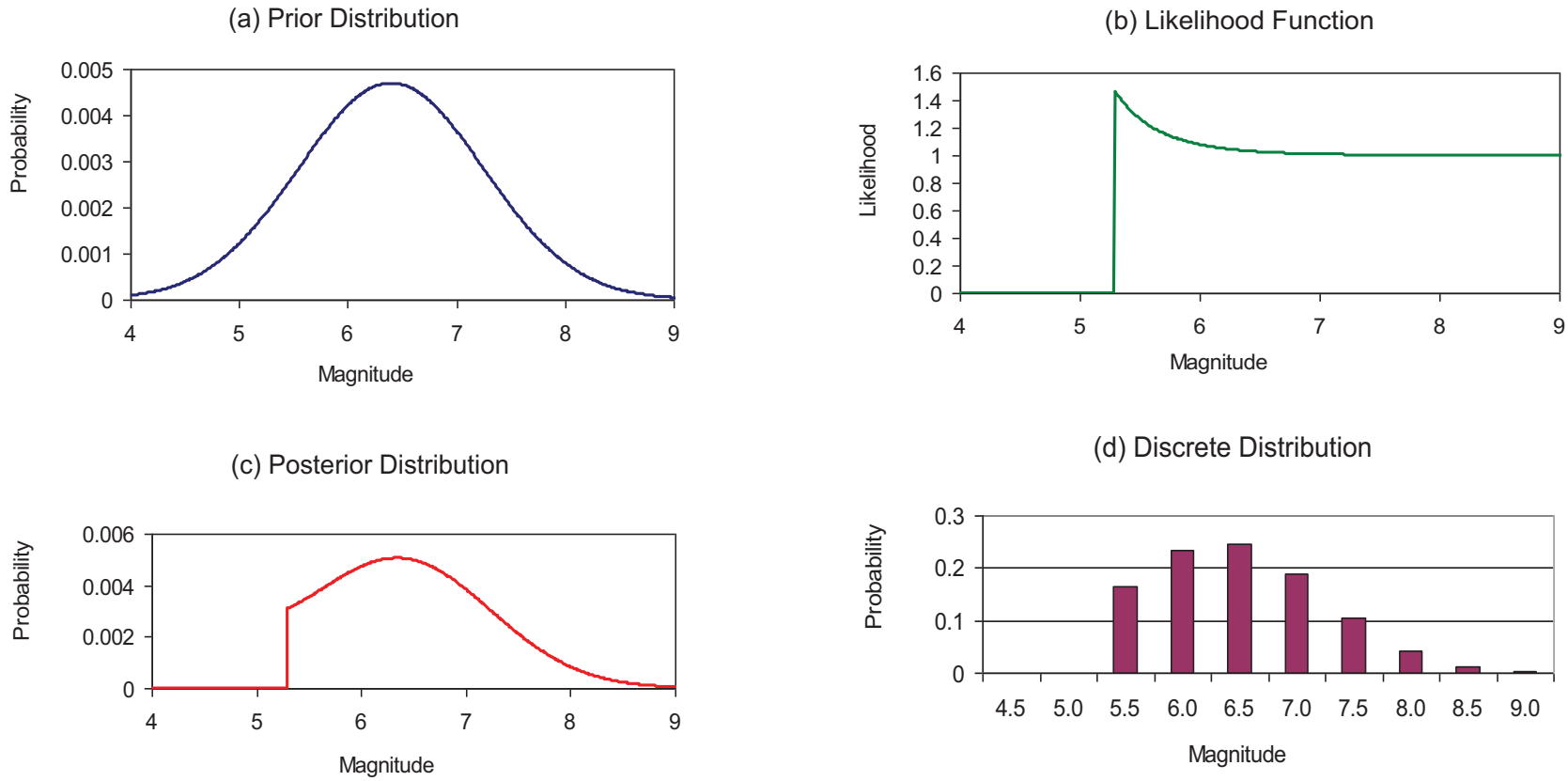

Figure 5.2.1-1
Diagrammatic illustration of the Bayesian Mmax approach showing (a) the prior distribution, (b) the likelihood function, and (c) the posterior distribution. The posterior distribution is represented by a discrete distribution (d) for implementation in hazard analysis.

Notes: Prior is normal distribution with mean M 6.4 and standard deviation 0.85. Likelihood function is based on two observed earthquakes larger than M 4.5, with the largest being 5.3.

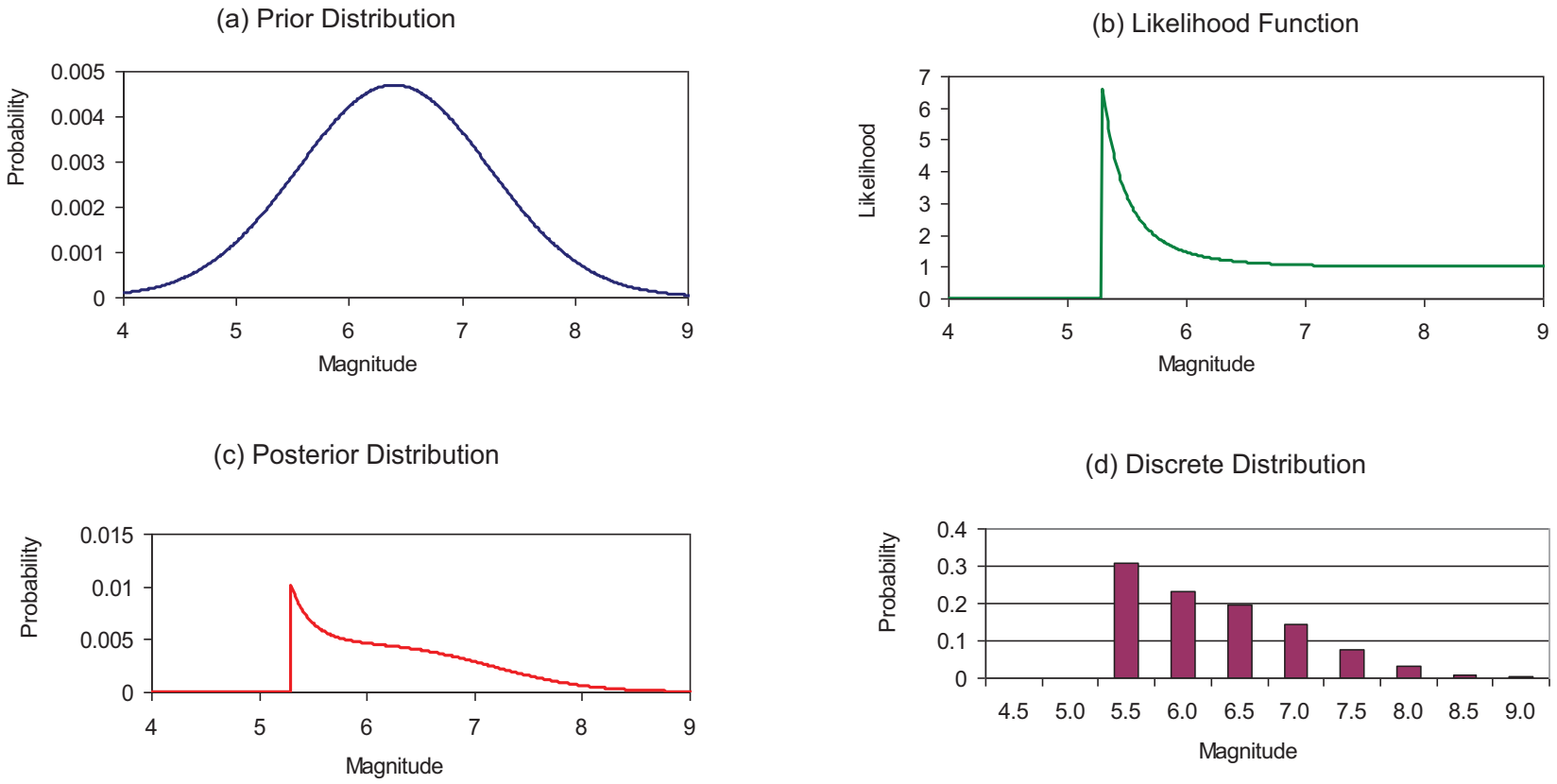

Figure 5.2.1-2
Diagrammatic illustration of the Bayesian Mmax approach showing (a) the prior distribution, (b) the likelihood function, and (c) the posterior distribution. The posterior distribution is represented by a discrete distribution (d) for implementation in hazard analysis.

Notes: Prior is normal distribution with mean M 6.4 and standard deviation 0.85. Likelihood function is based on 10 observed earthquakes larger than M 4.5, with the largest being 5.3.

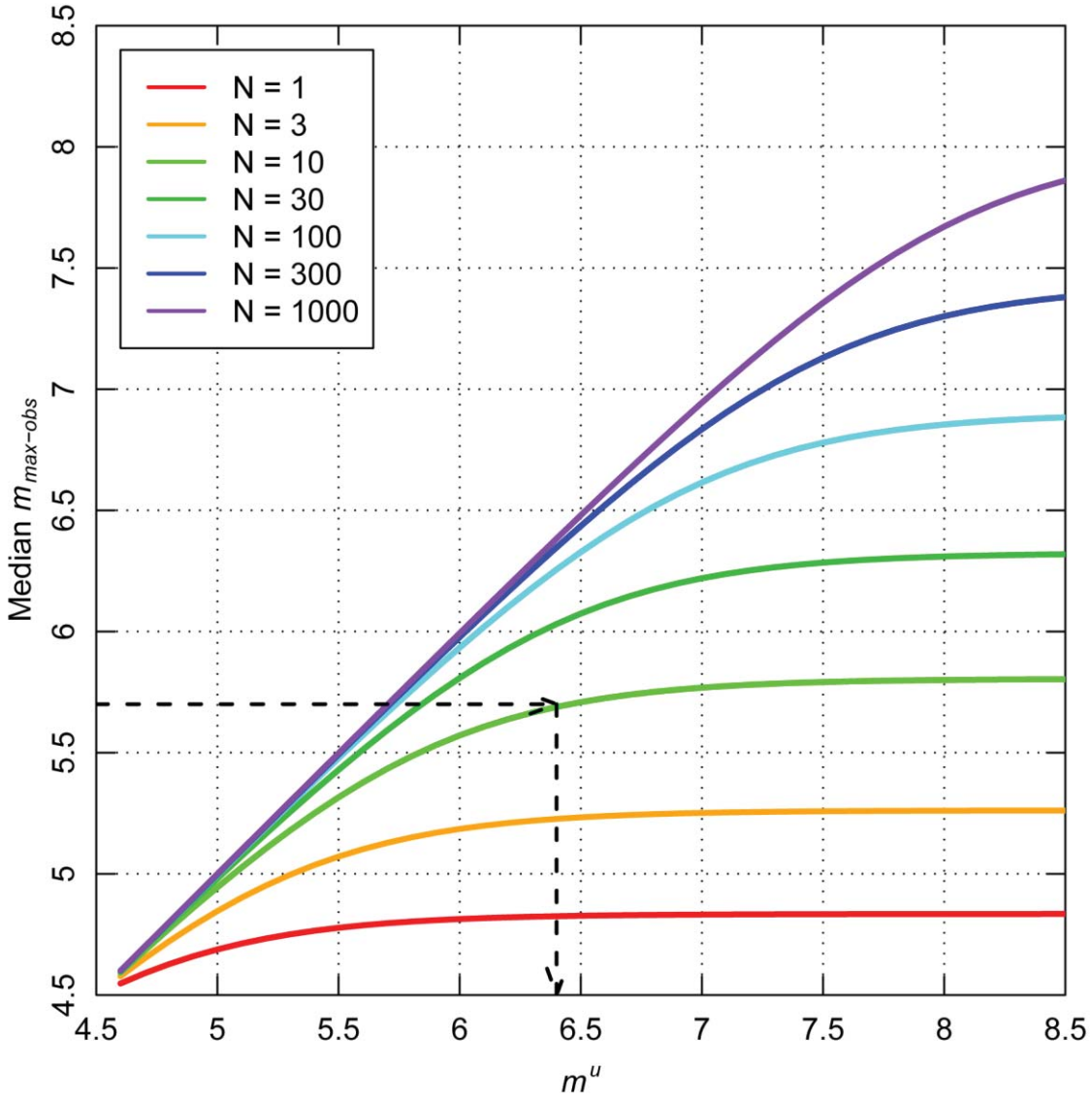

Figure 5.2.1-3
Median values of $m_{\max-obs}$ as a function of maximum magnitude, m^u, and sample size *N*, the number of earthquakes ≥ M 4.5

Note: Dashed lines show bias correction for a $m_{\max-obs}$ of 5.7 with a sample size of 10.

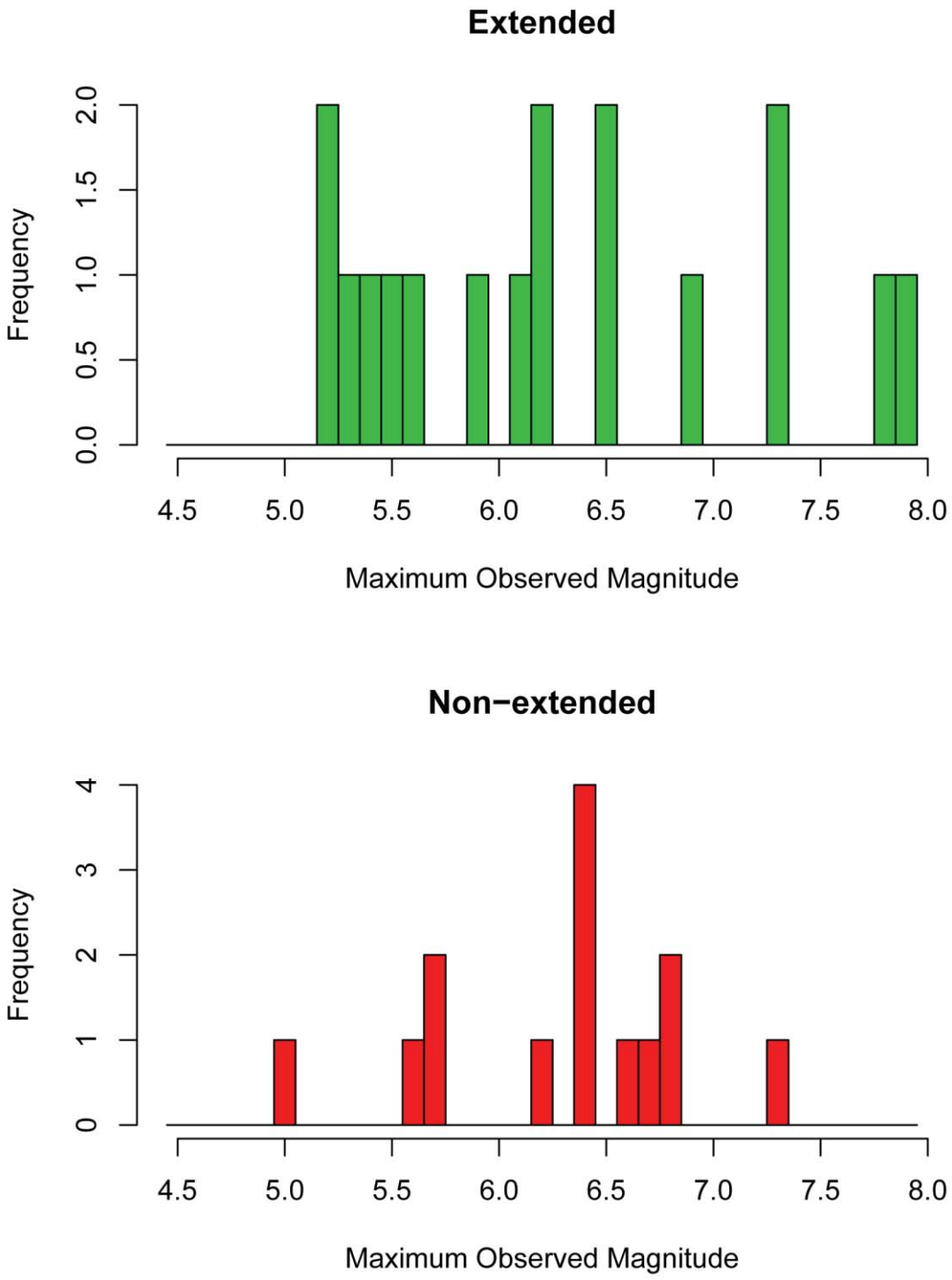

Figure 5.2.1-4
Histograms of $m_{\max-obs}$ for extended and non-extended superdomains

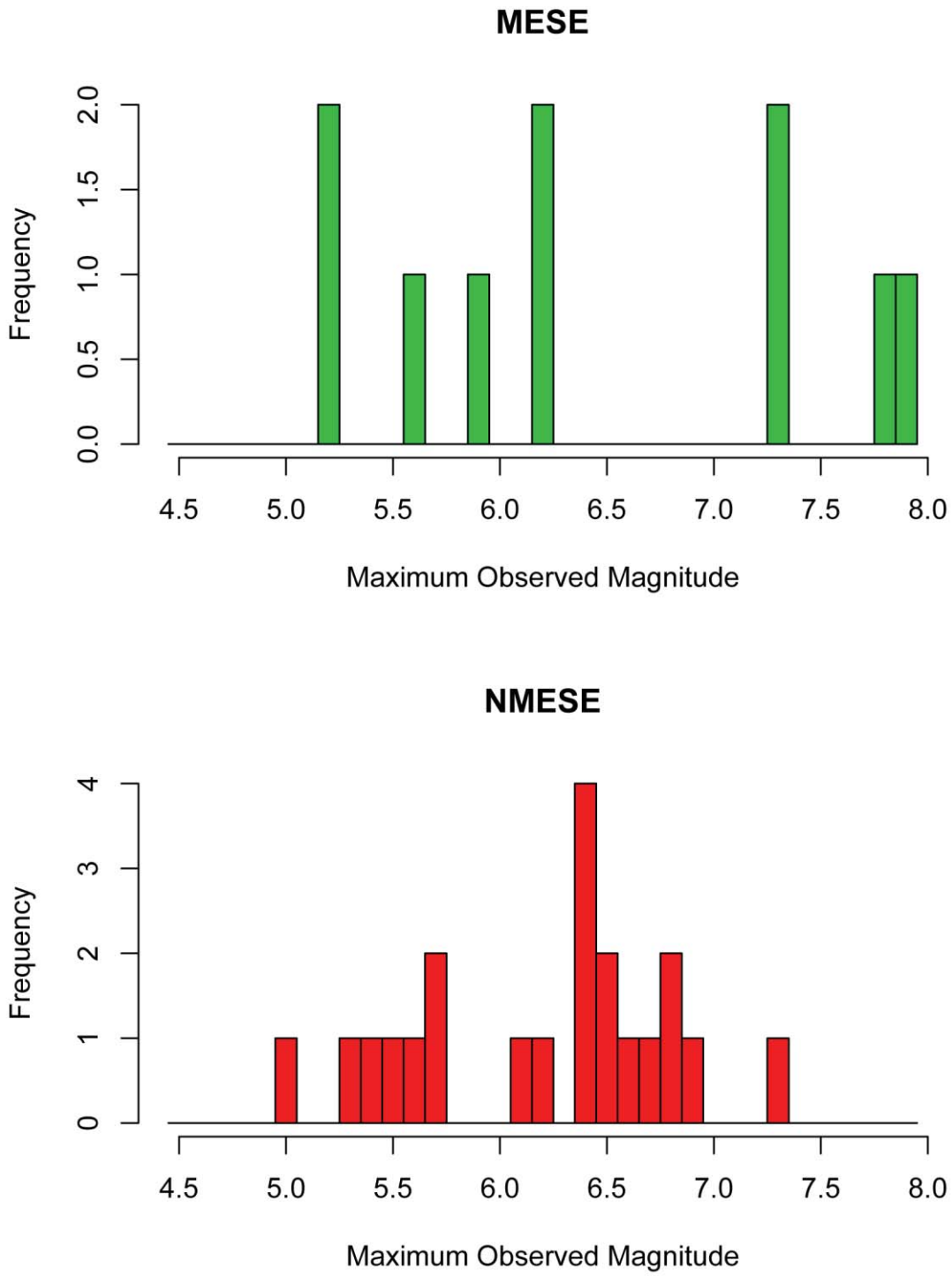

Figure 5.2.1-5
Histograms of $m_{\max-obs}$ for Mesozoic-and-younger extended (MESE) superdomains and for older extended and non-extended (NMESE) superdomains

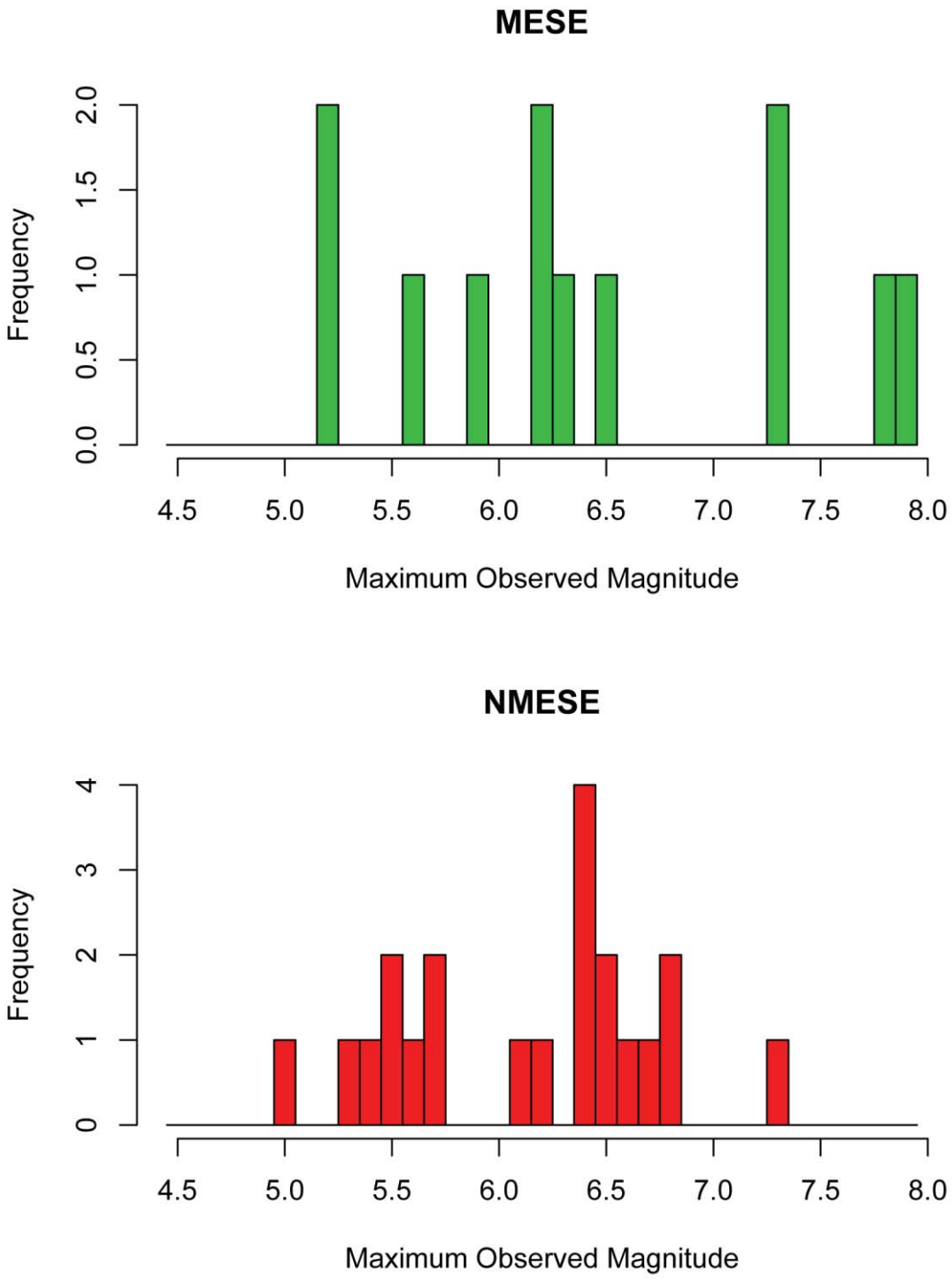

Figure 5.2.1-6
Histograms of $m_{\max-obs}$ for Mesozoic-and-younger extended (MESE) superdomains and for older extended and non-extended (NMESE) superdomains using age of most recent extension for the age classification

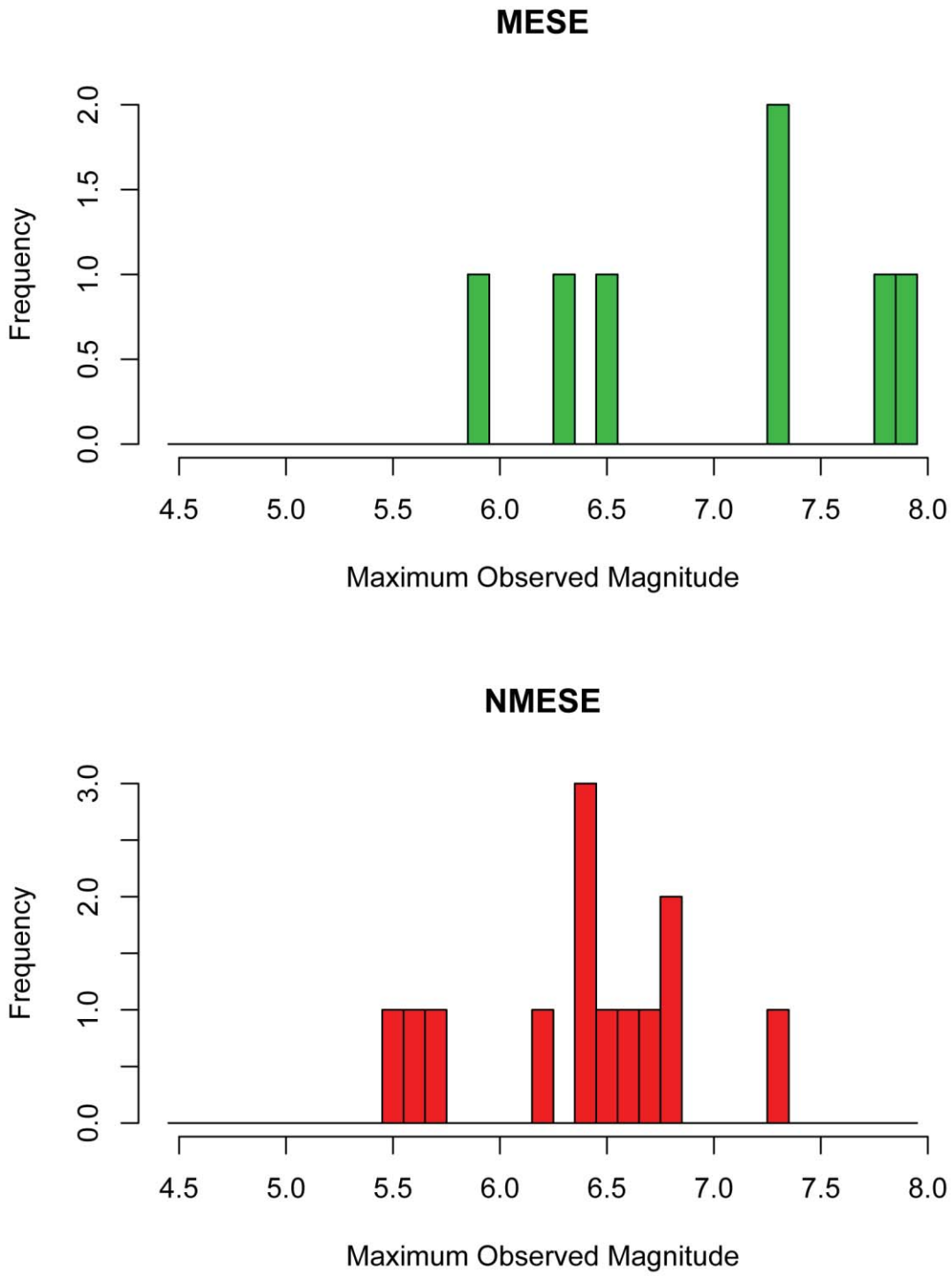

Figure 5.2.1-7
Histograms of $m_{\max-obs}$ for Mesozoic-and-younger extended (MESE) superdomains and for older extended and non-extended (NMESE) superdomains using final sets indicated by asterisks in Tables 5.2.1-1 and 5.2.1-2

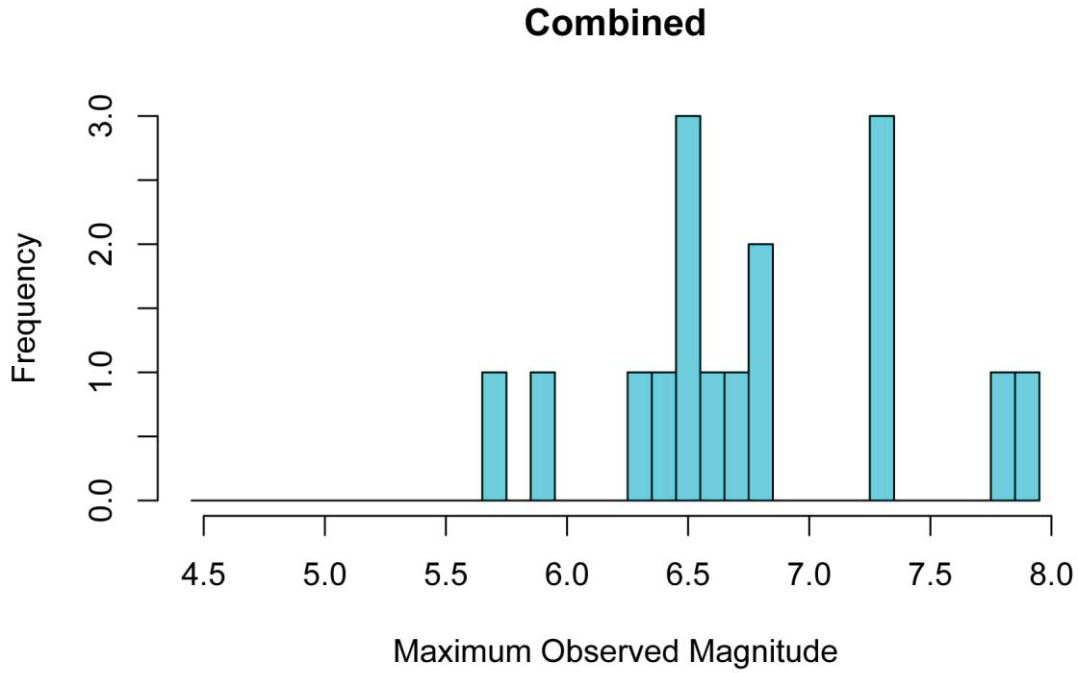

Figure 5.2.1-8
Histograms of $m_{\max-obs}$ for combined (COMB) superdomains using final sets indicated by asterisks in Table 5.2.1-3

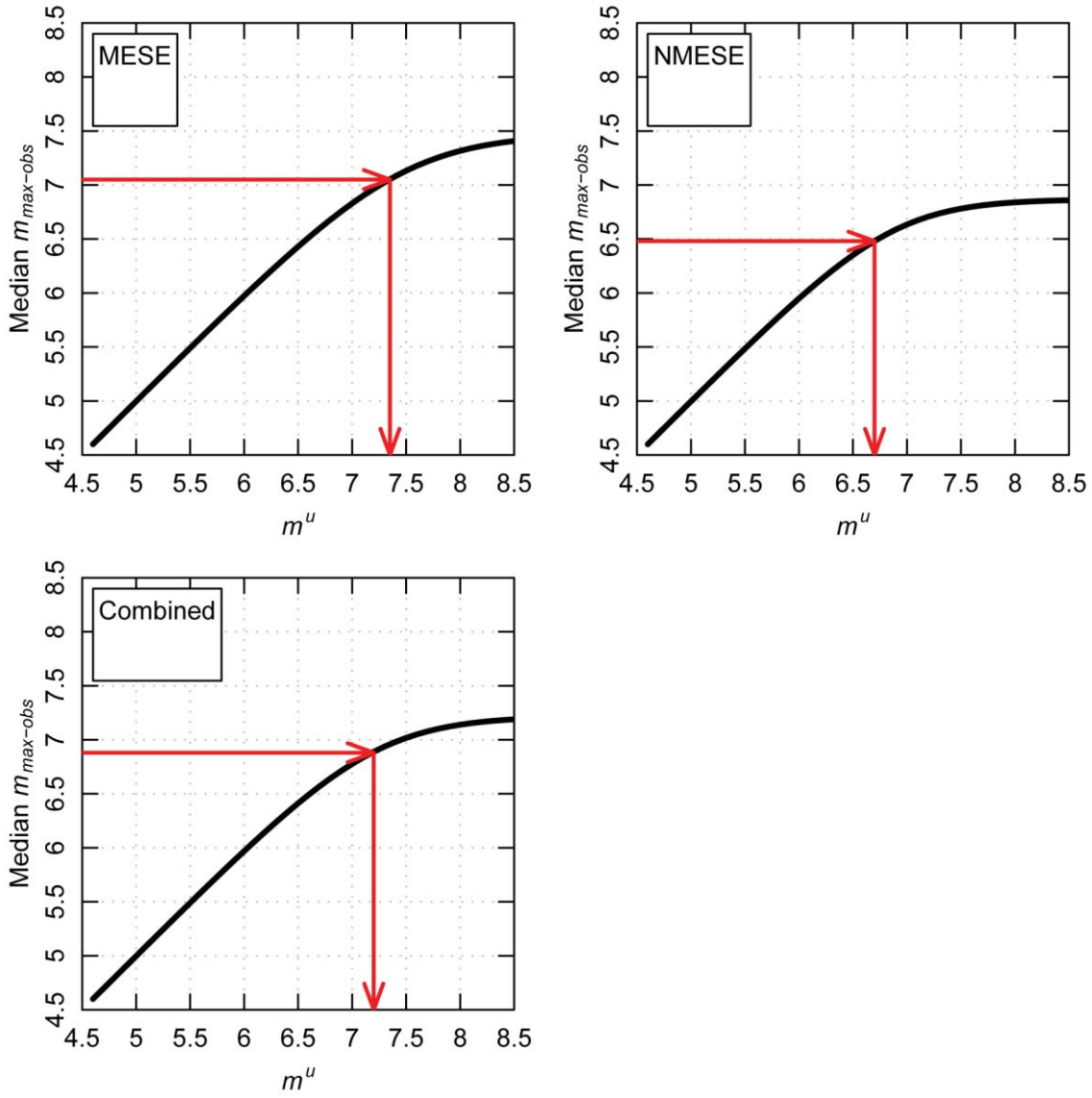

Figure 5.2.1-9
Bias adjustments from $m_{\max-obs}$ to m^u for the three sets of superdomain analysis results presented in Table 5.2.1-4

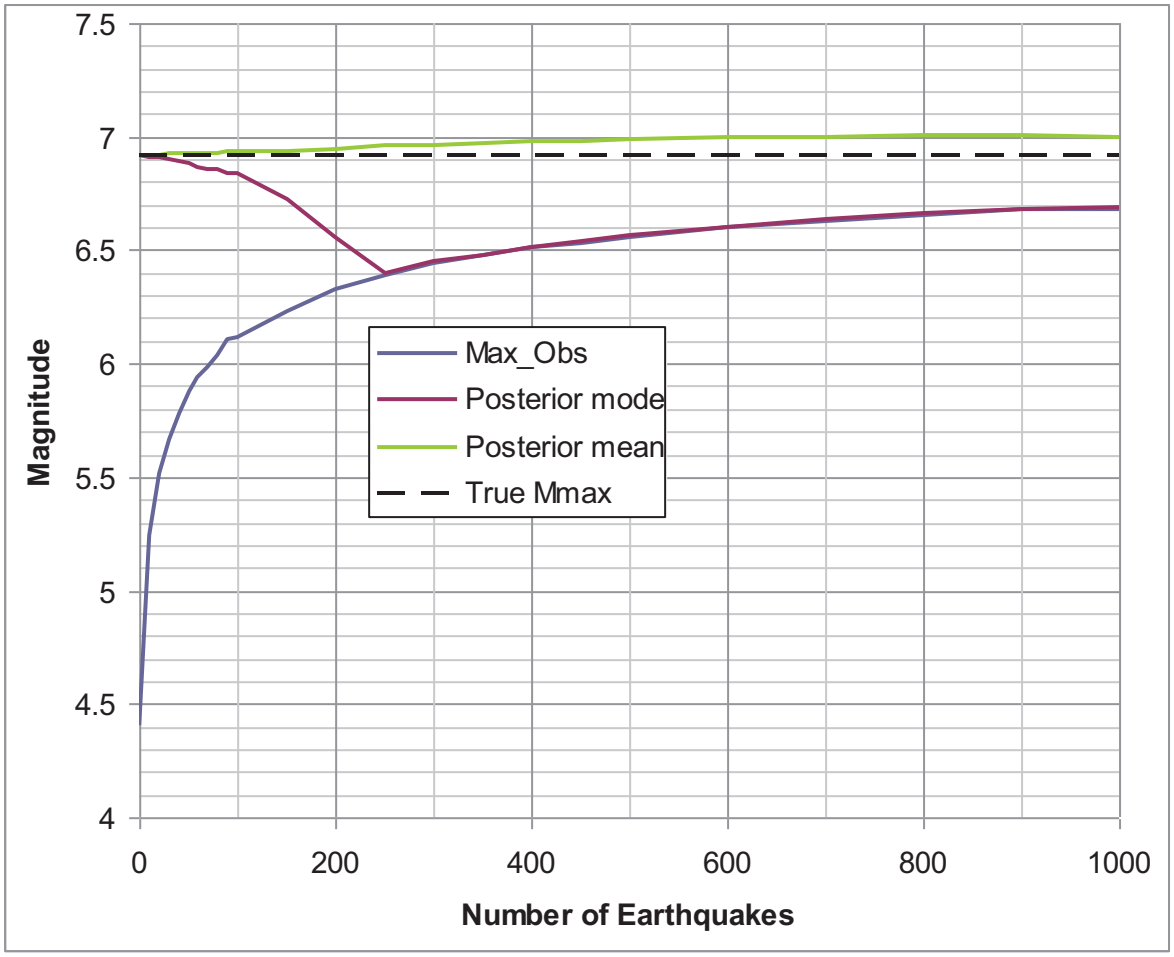

Figure 5.2.1-10
Results of simulations of estimates of Mmax using the Bayesian approach for earthquake catalogs ranging in size from 1 to 1,000 earthquakes. True Mmax is set at the mean of the prior distribution.

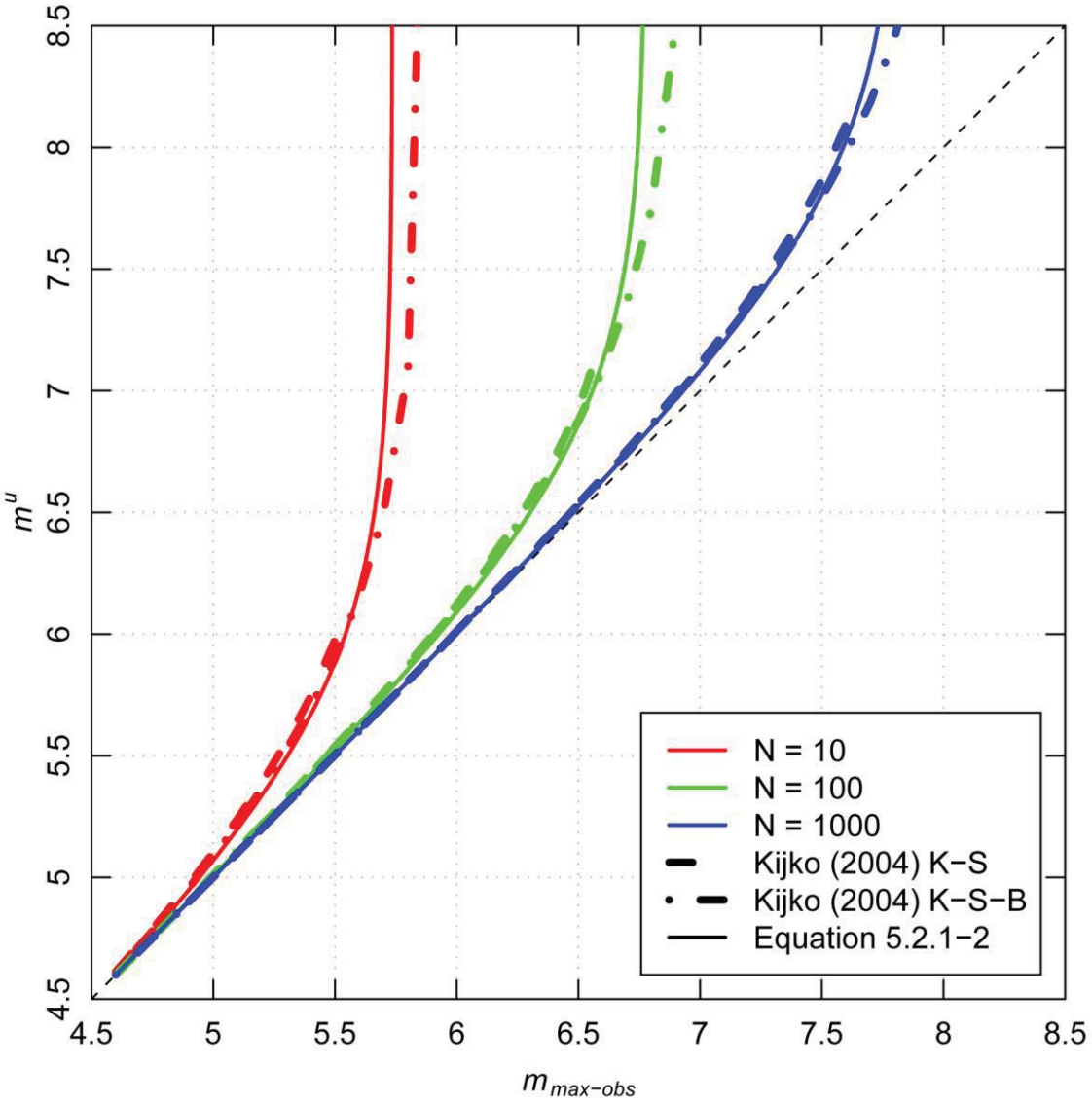

Figure 5.2.1-11
Comparison of the Kijko (2004) estimates of m^u for given values of $m_{\max-obs}$ and *N*, the number of earthquakes of magnitude ≥ 4.5. Also shown is the median value of $m_{\max-obs}$ for given m^u obtained using Equation 5.2.1-2.

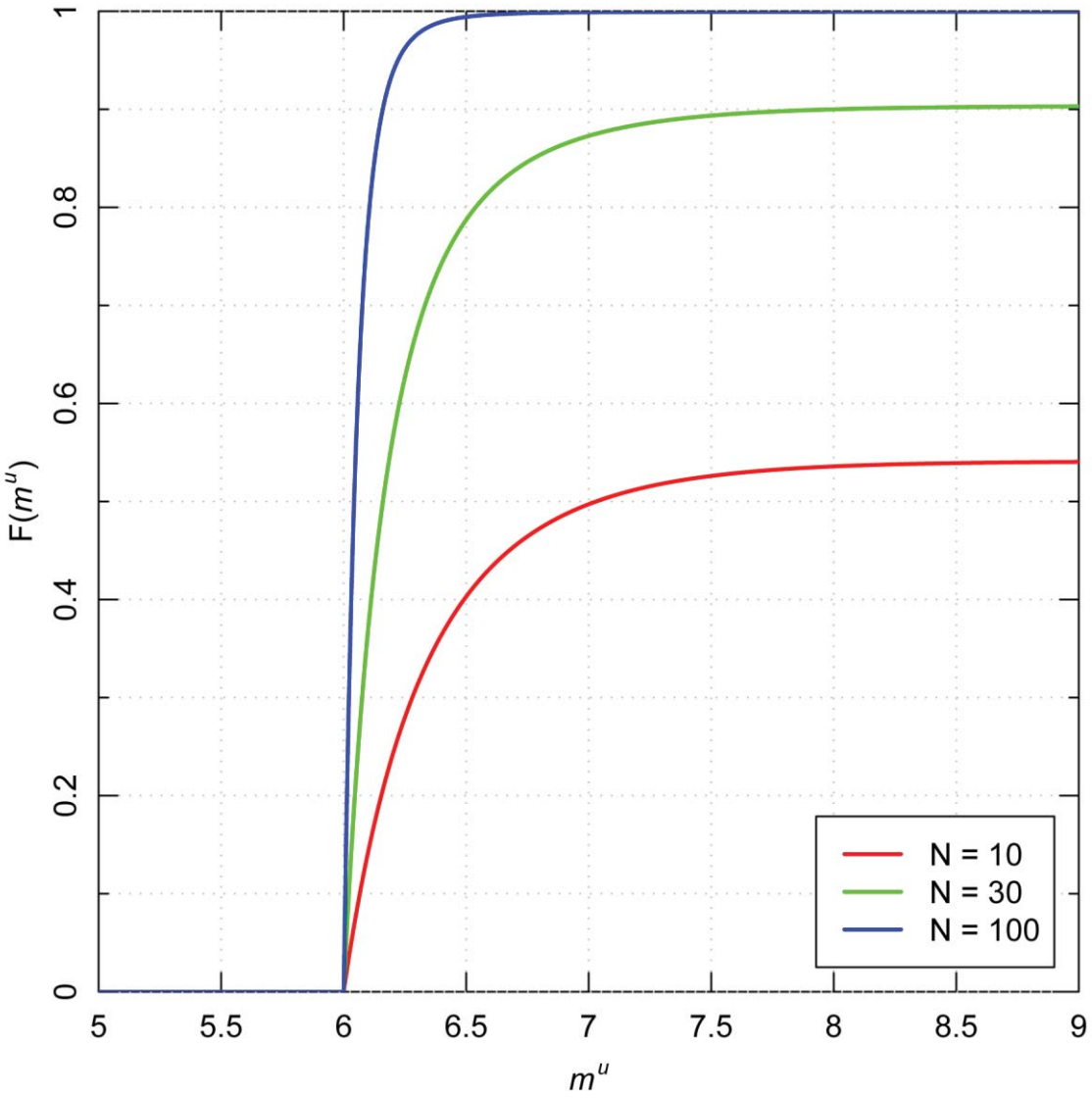

Figure 5.2.1-12
Behavior of the cumulative probability function for m^u (Equation 5.2.1-9) for the K-S-B estimator and a value of $m_{\max-obs}$ equal to 6

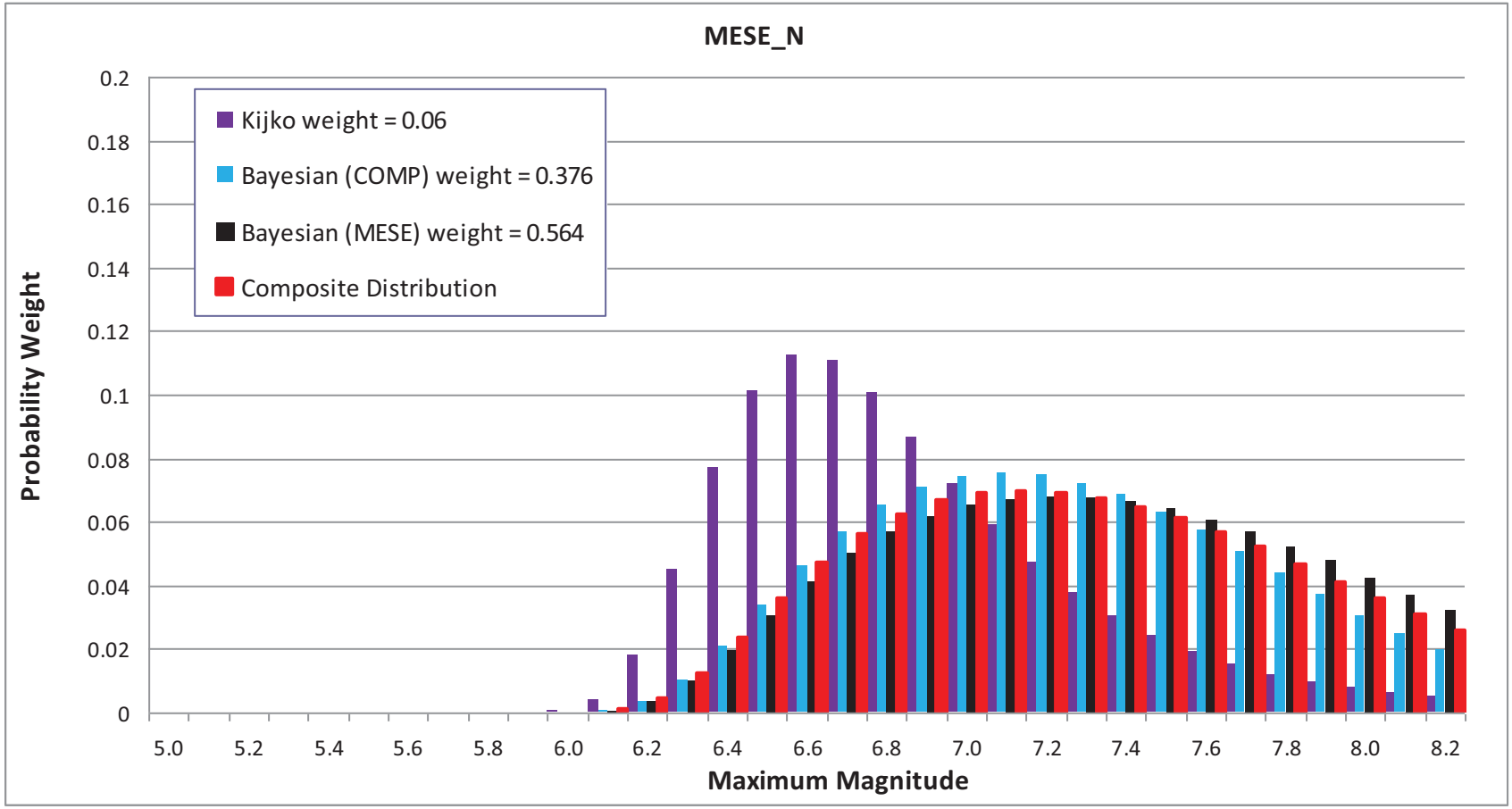

Figure 5.2.1-13
Example Mmax distribution assessed for the Mesozoic-and-younger extended Mmax zone for the case where the zone is "narrow" (MESE-N). Distributions are shown for the Kijko approach and for the Bayesian approach using either the Mesozoic-and-younger extended prior distribution or the composite prior distribution. The final composite Mmax distribution, which incorporates the relative weights, is shown by the red probability distribution.

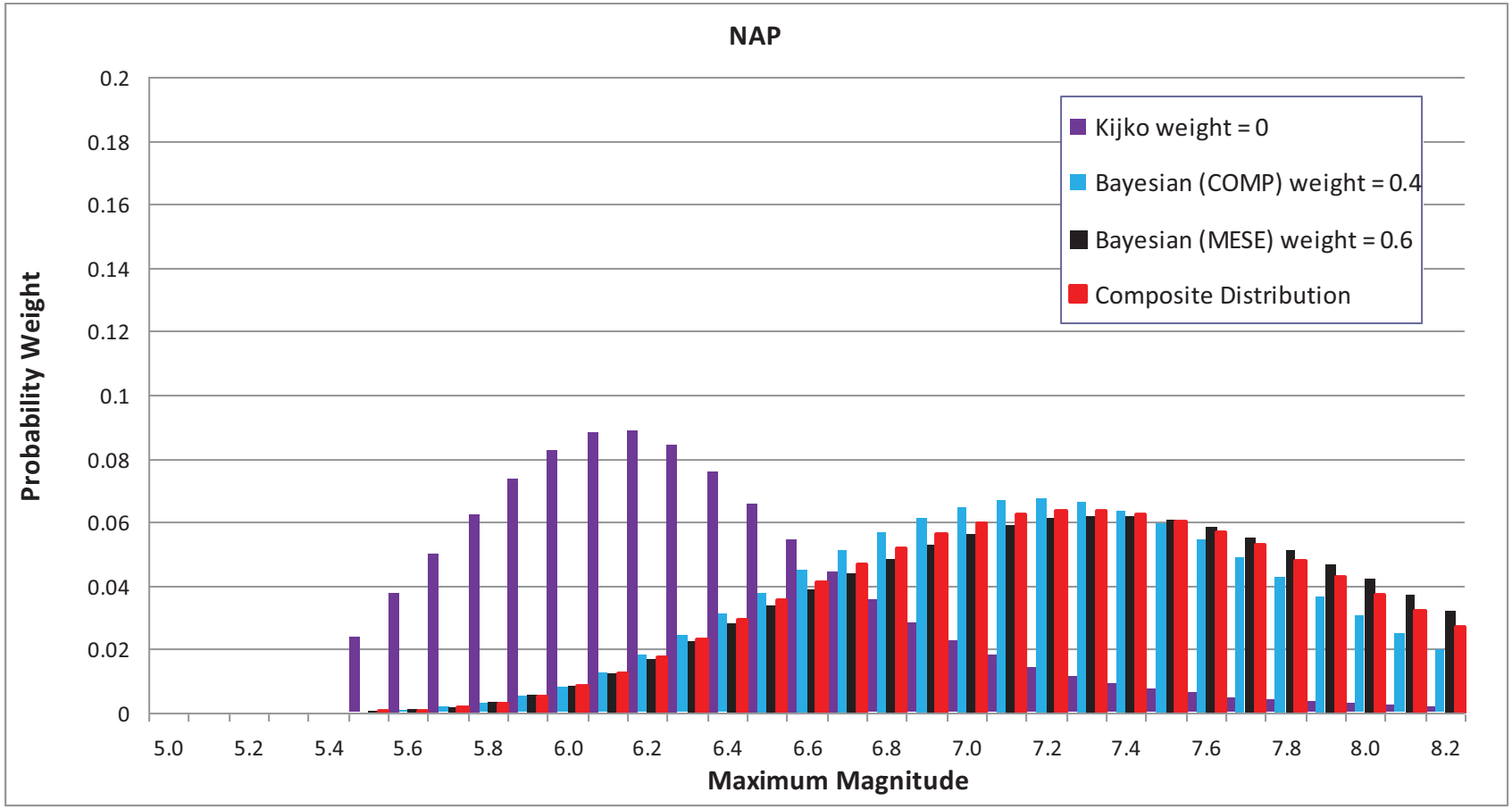

Figure 5.2.1-14
Example Mmax distribution assessed for the Northern Appalachian seismotectonic zone (NAP). Distributions are shown for the Kijko approach and for the Bayesian approach using either the Mesozoic-and-younger extended prior distribution or the composite prior distribution. Note that the Kijko results are shown in this example for illustration, even though they have zero weight. The final composite Mmax distribution, which incorporates the relative weights, is shown by the red probability distribution.

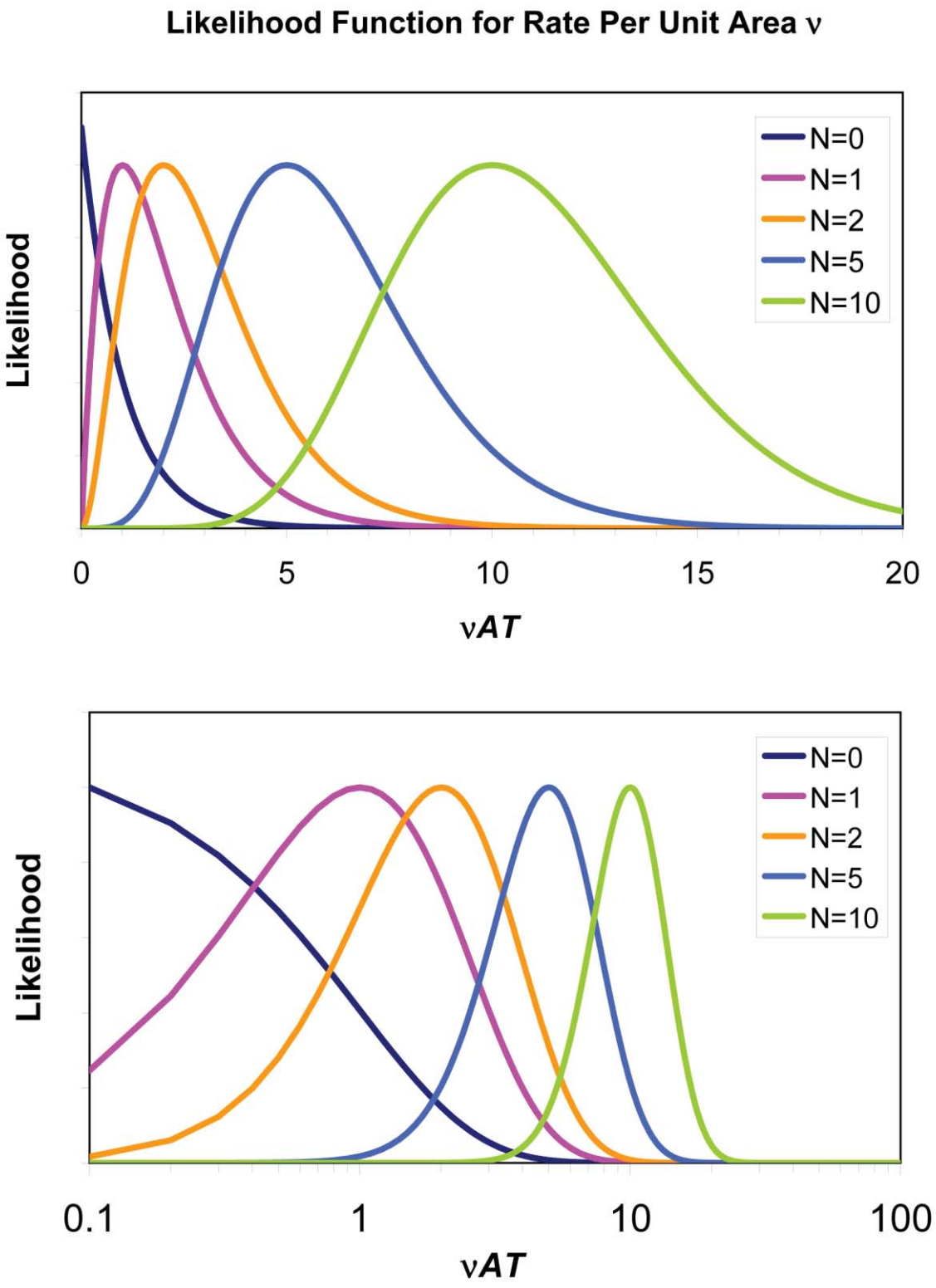

Figure 5.3.2-1
Likelihood function for rate per unit area in a Poisson process, for multiple values of the earthquake count *N*: (a) arithmetic scale, and (b) logarithmic scale used to illustrate decreasing COV as N increases

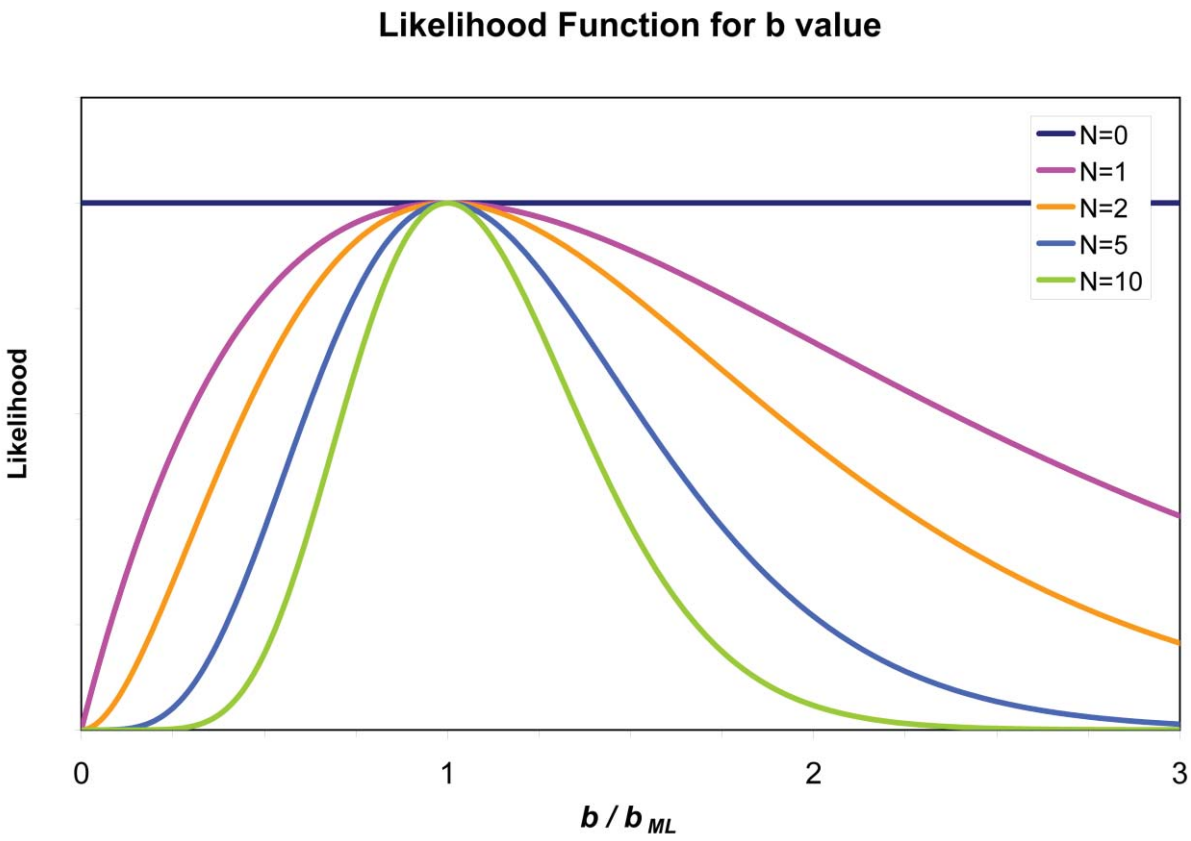

Figure 5.3.2-2
Likelihood function for *b*-value of an exponential magnitude distribution, for multiple values of the earthquake count *N*. The value of *b* is normalized by the maximum-likelihood estimate, which is derived from Equation 5.3.2-5.

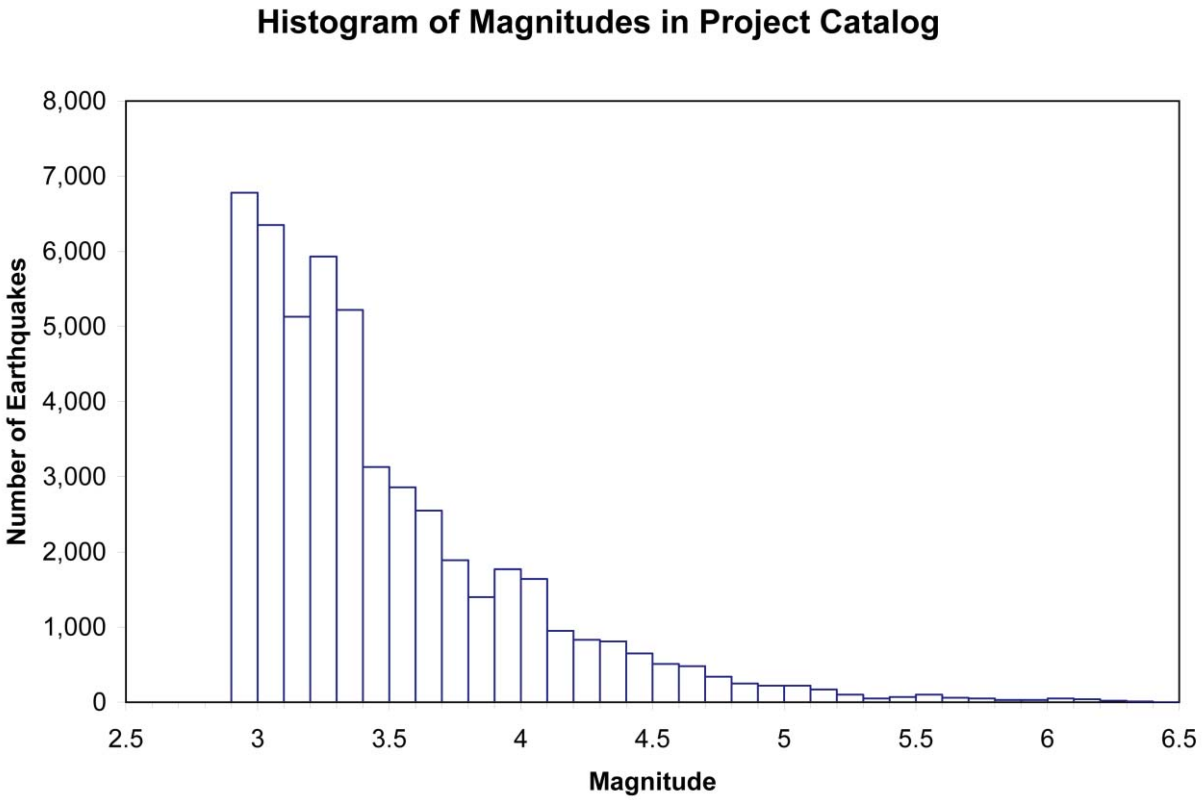

Figure 5.3.2-3
Histogram of magnitudes in the earthquake catalog used in this section. The minimum magnitude shown (M 2.9) is the lowest magnitude used in these recurrence calculations.

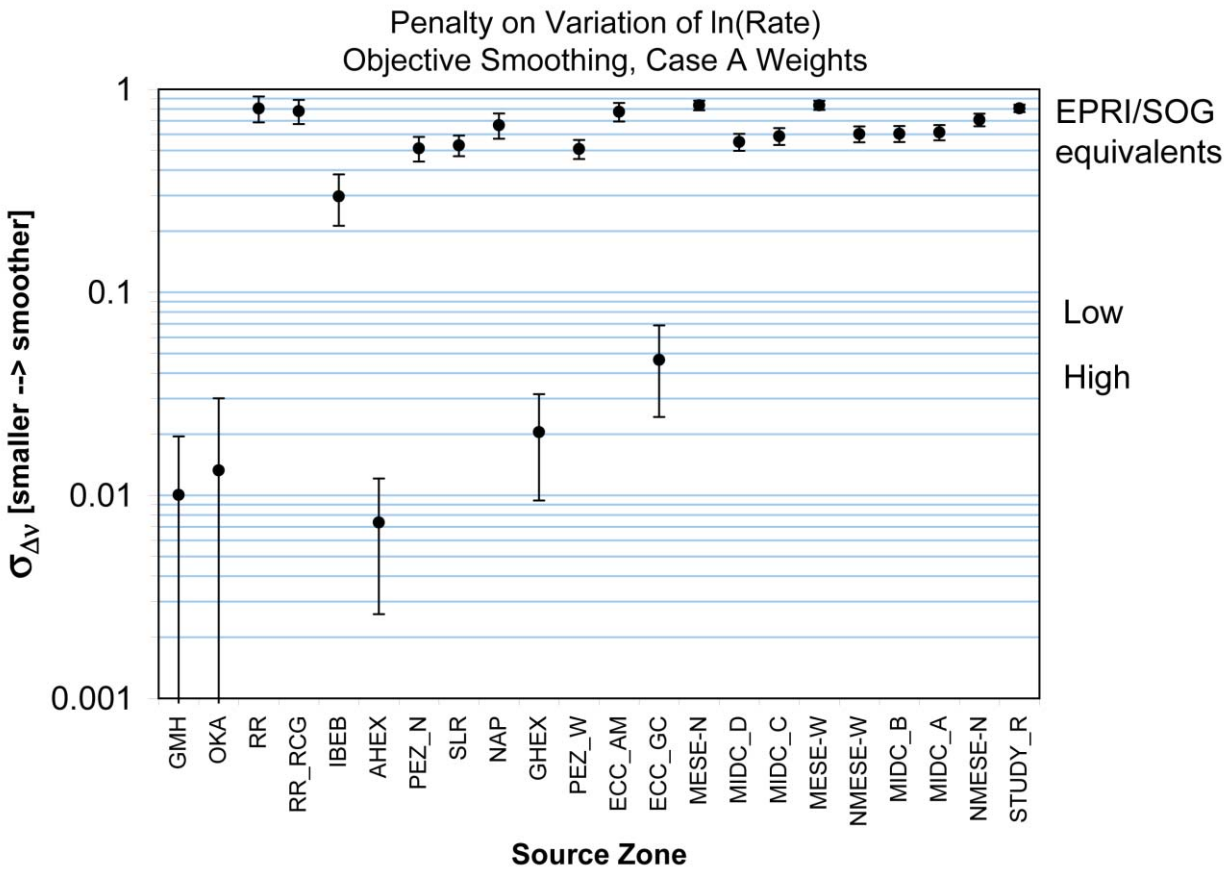

Figure 5.3.2-4
Objectively determined values of the penalty function for ln(rate) for Case A magnitude weights. Source zones are sorted from smallest to largest. See list of abbreviations for full source-zone names.

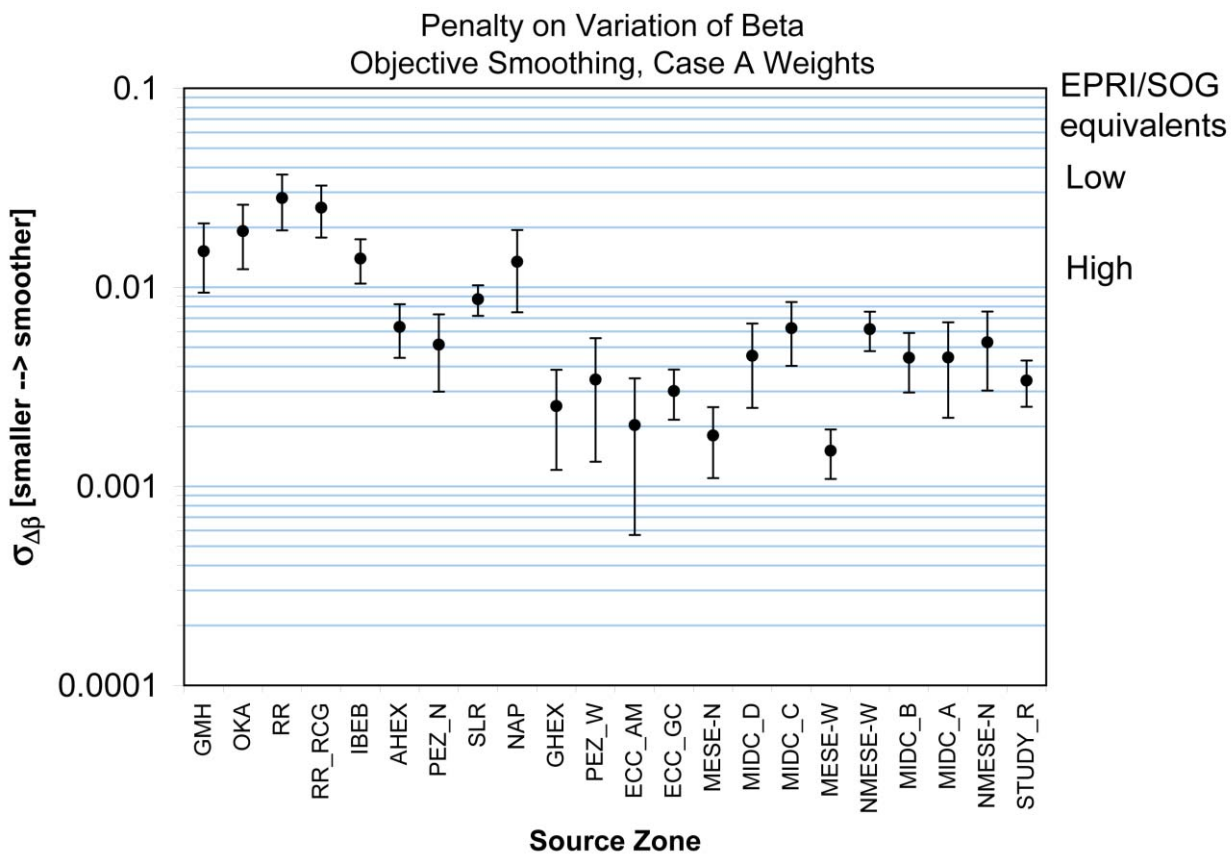

Figure 5.3.2-5
Objectively determined values of the penalty function for beta for Case A magnitude weights

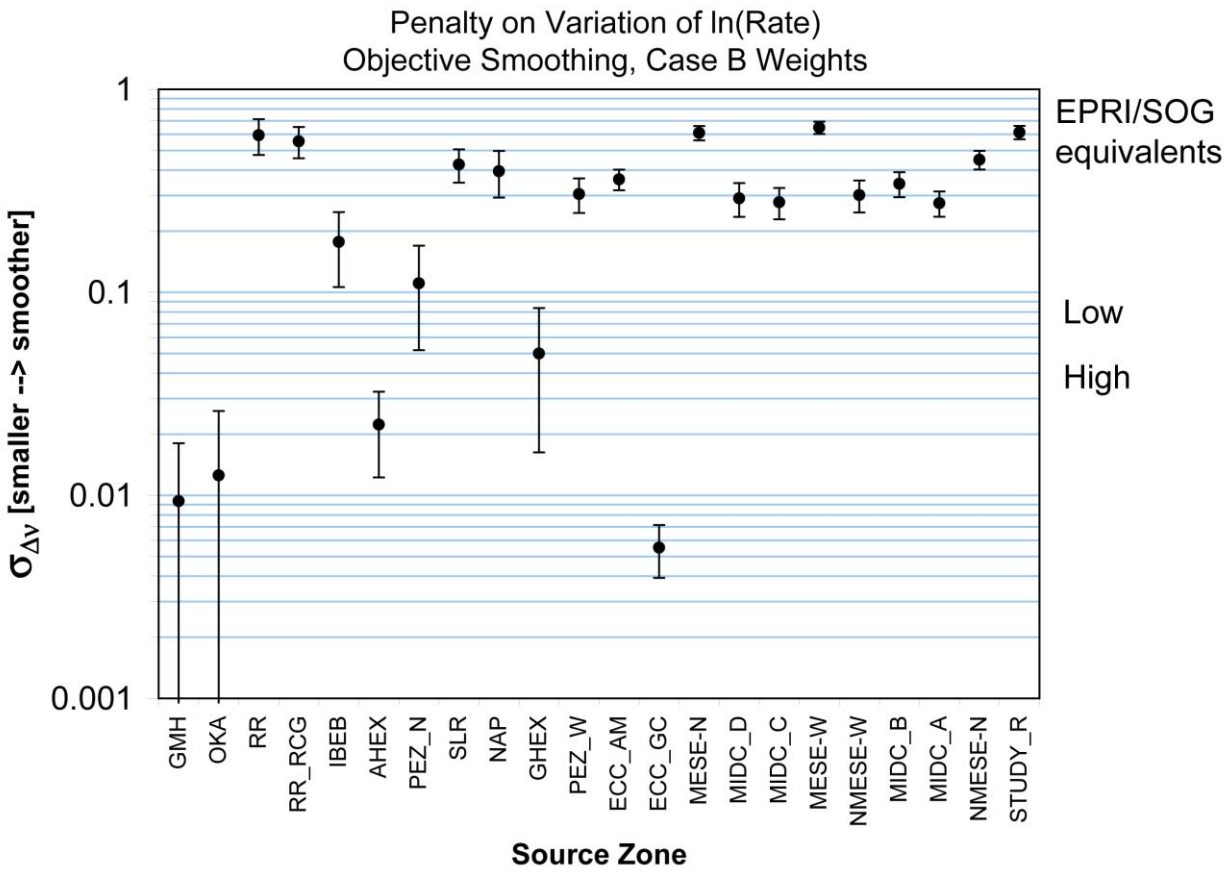

Figure 5.3.2-6
Objectively determined values of the penalty function for ln(rate) for Case B magnitude weights

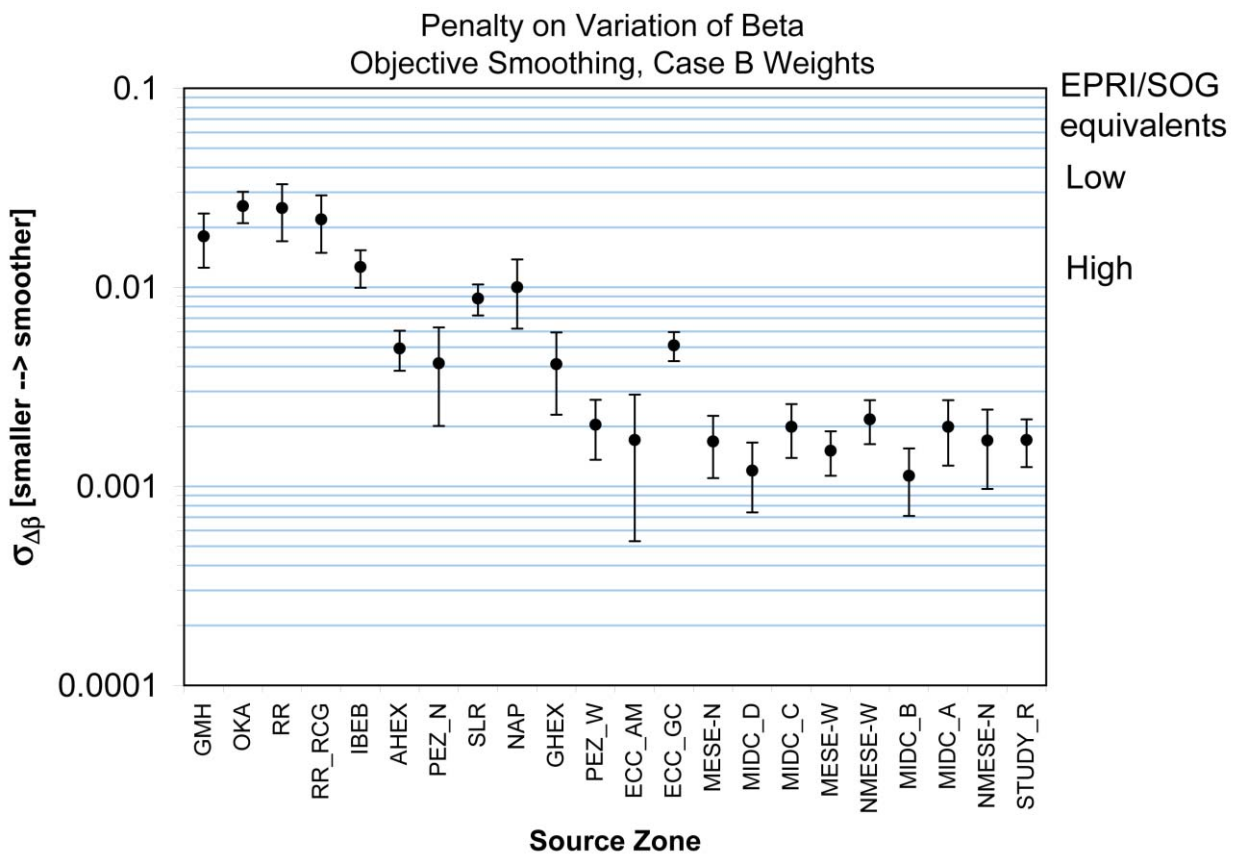

Figure 5.3.2-7
Objectively determined values of the penalty function for beta for Case B magnitude weights. Source zones are sorted from smallest to largest.

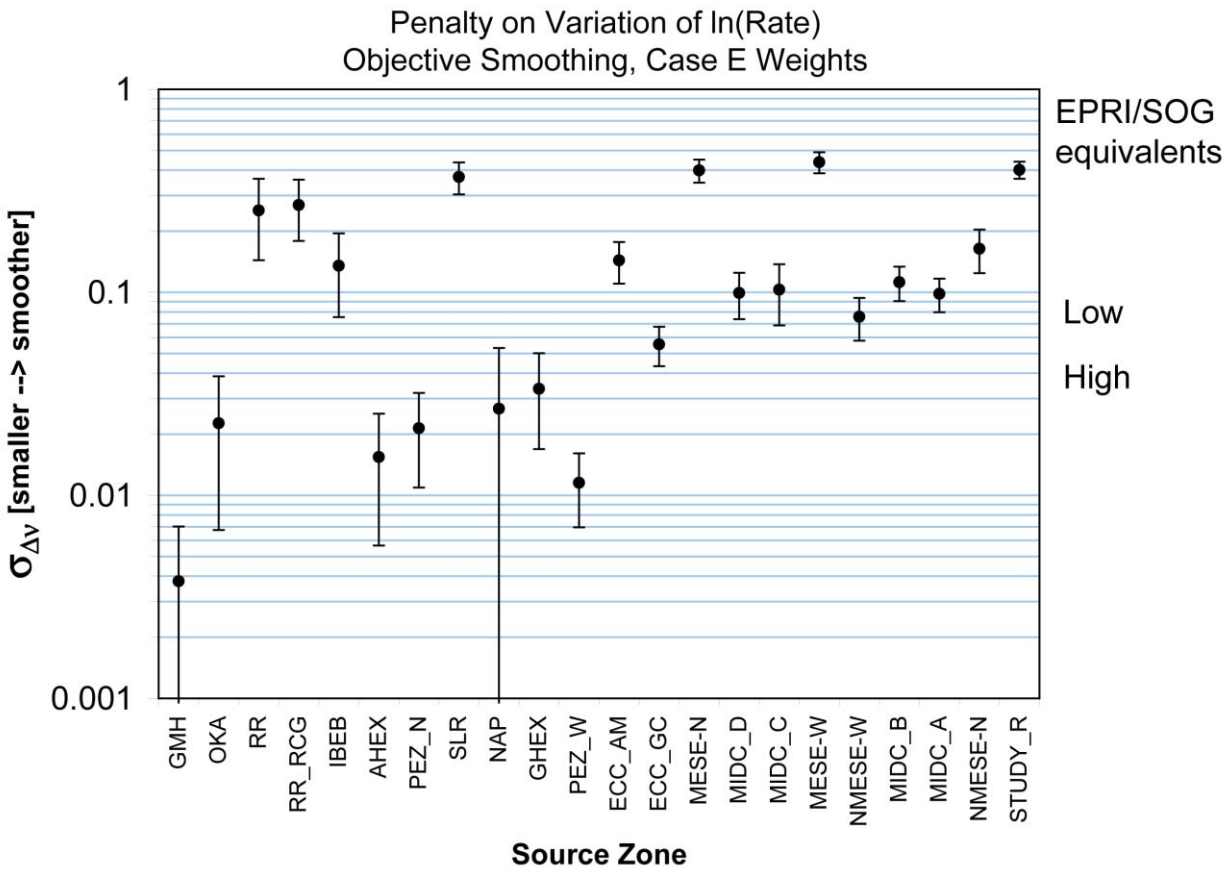

Figure 5.3.2-8
Objectively determined values of the penalty function for ln(rate) for Case E magnitude weights

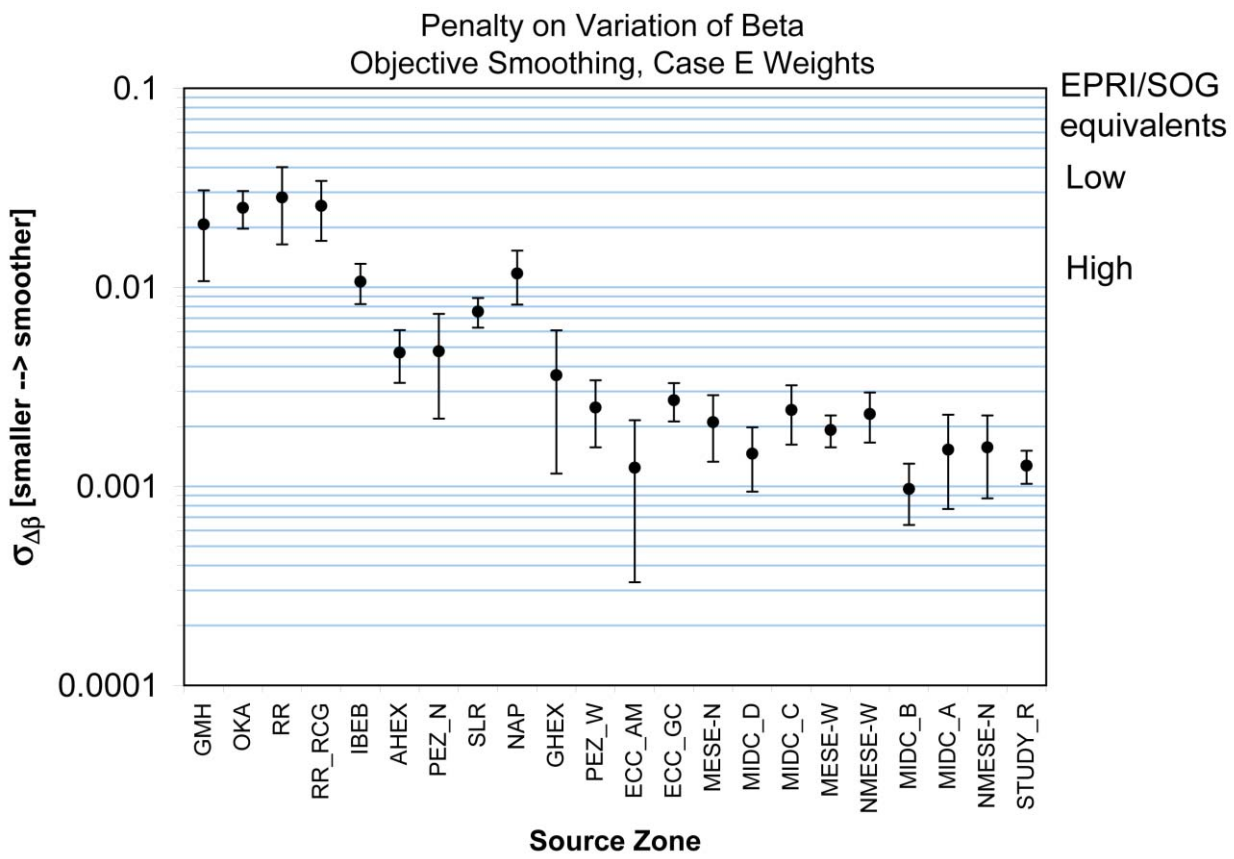

Figure 5.3.2-9
Objectively determined values of the penalty function for beta for Case E magnitude weights. Source zones are sorted from smallest to largest.

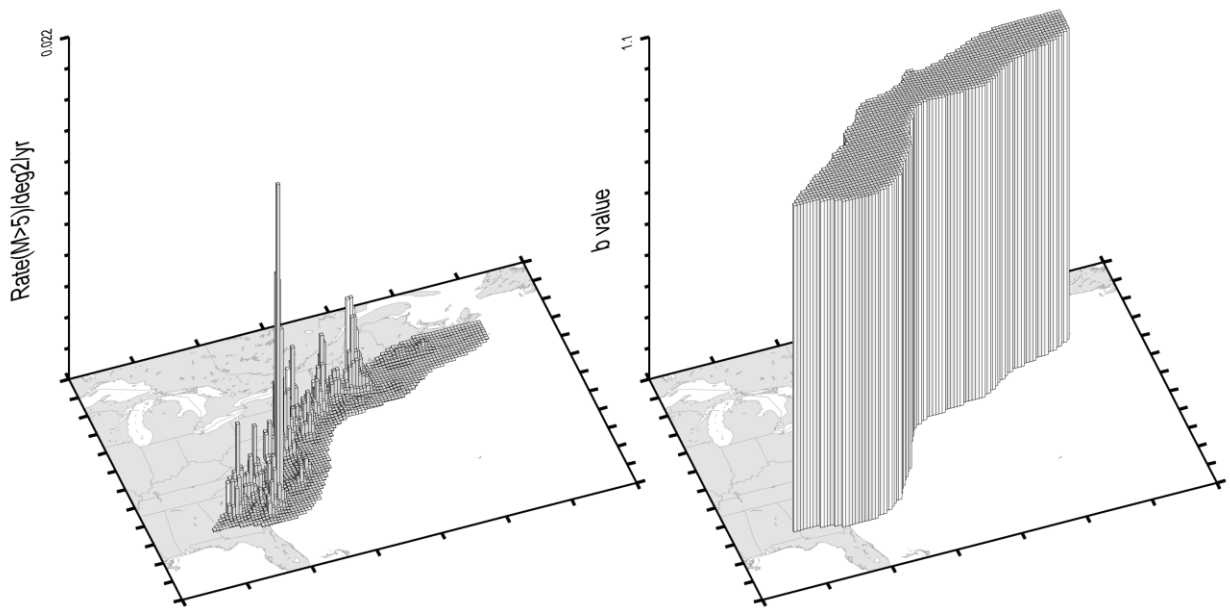

Figure 5.3.2-10
Mean map of rate and *b*-value for ECC-AM calculated using Case A magnitude weights

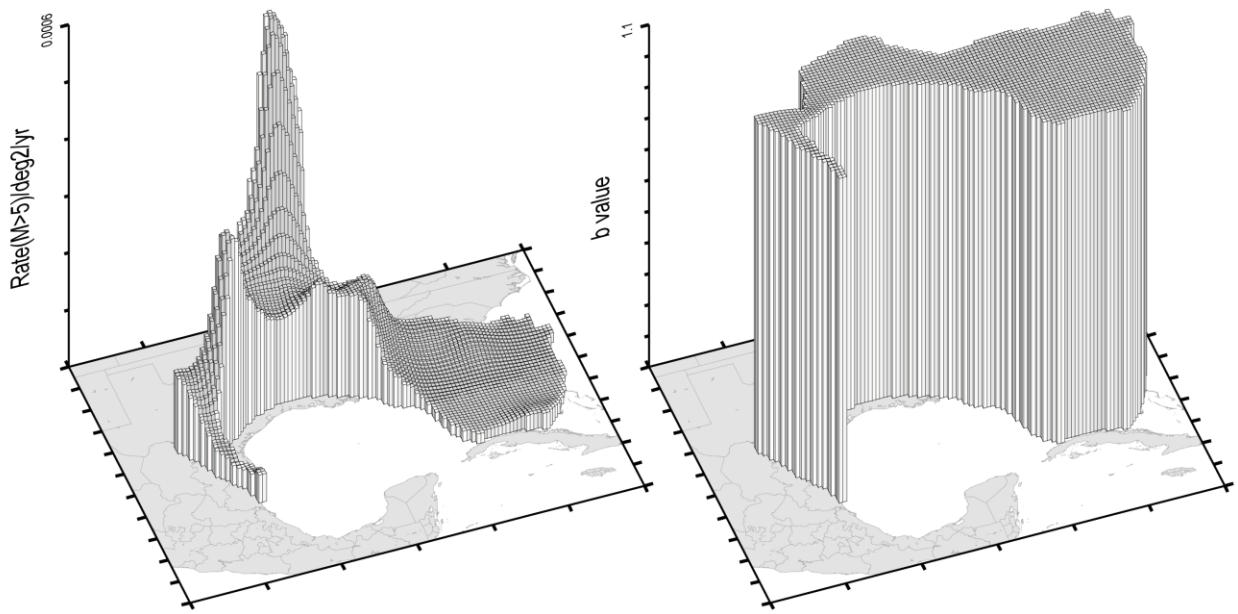

Figure 5.3.2-11
Mean map of rate and *b*-value for ECC-GC calculated using Case A magnitude weights

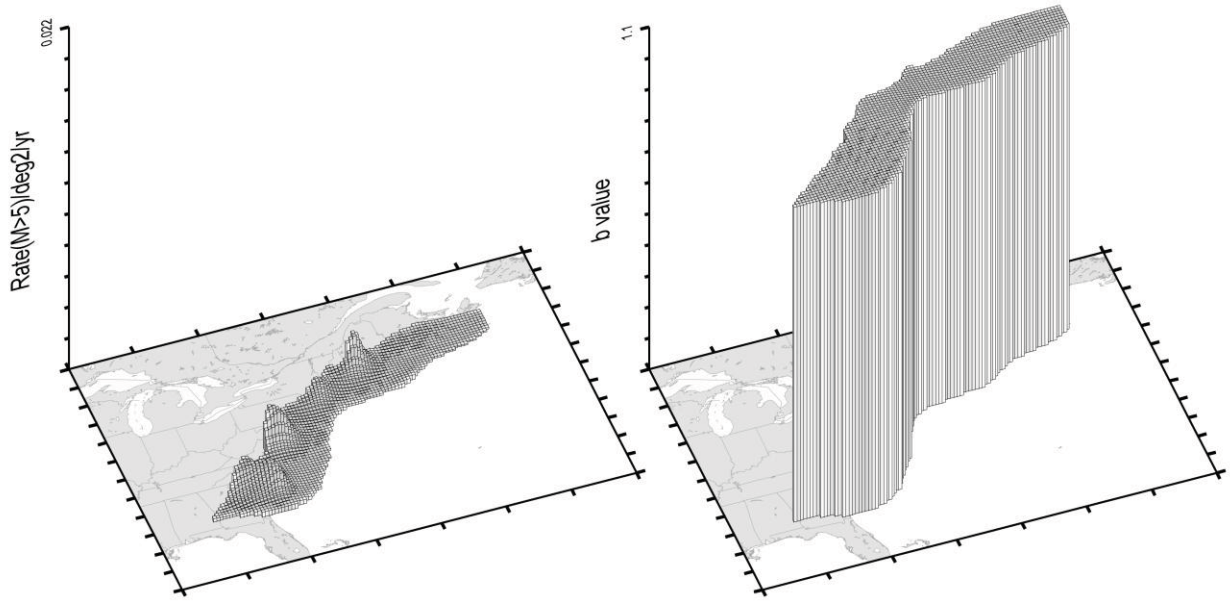

Figure 5.3.2-12
Mean map of rate and *b*-value for ECC-AM calculated using Case B magnitude weights

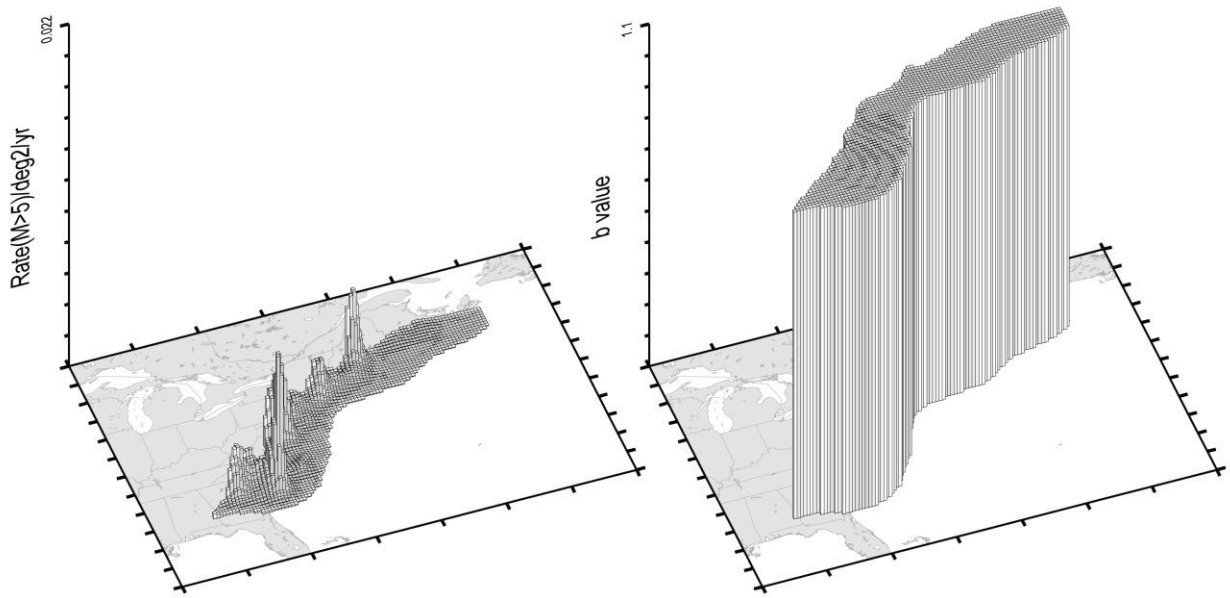

Figure 5.3.2-13
Mean map of rate and *b*-value for ECC-GC calculated using Case B magnitude weights

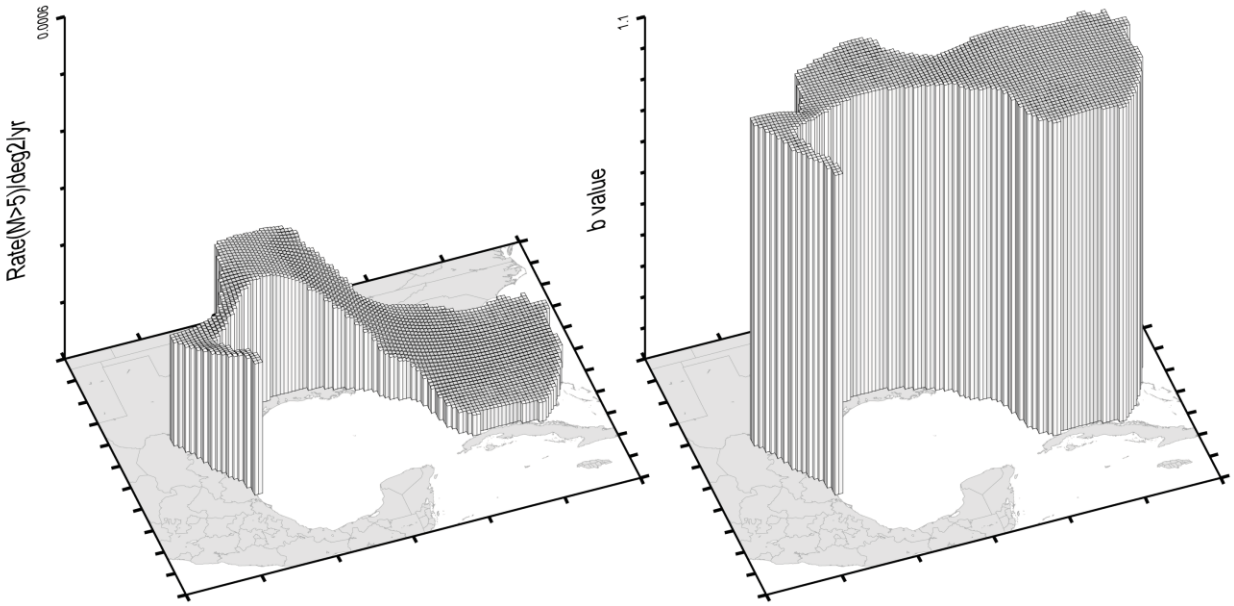

Figure 5.3.2-14
Mean map of rate and *b*-value for ECC-AM calculated using Case E magnitude weights

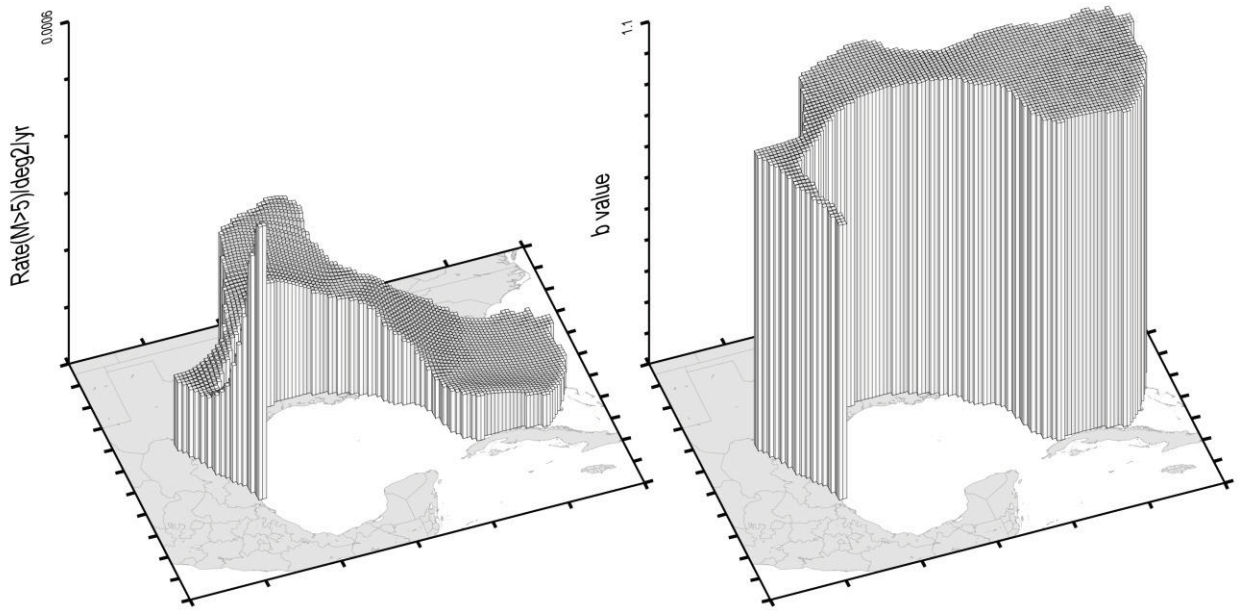

Figure 5.3.2-15
Mean map of rate and *b*-value for ECC-GC calculated using Case E magnitude weights

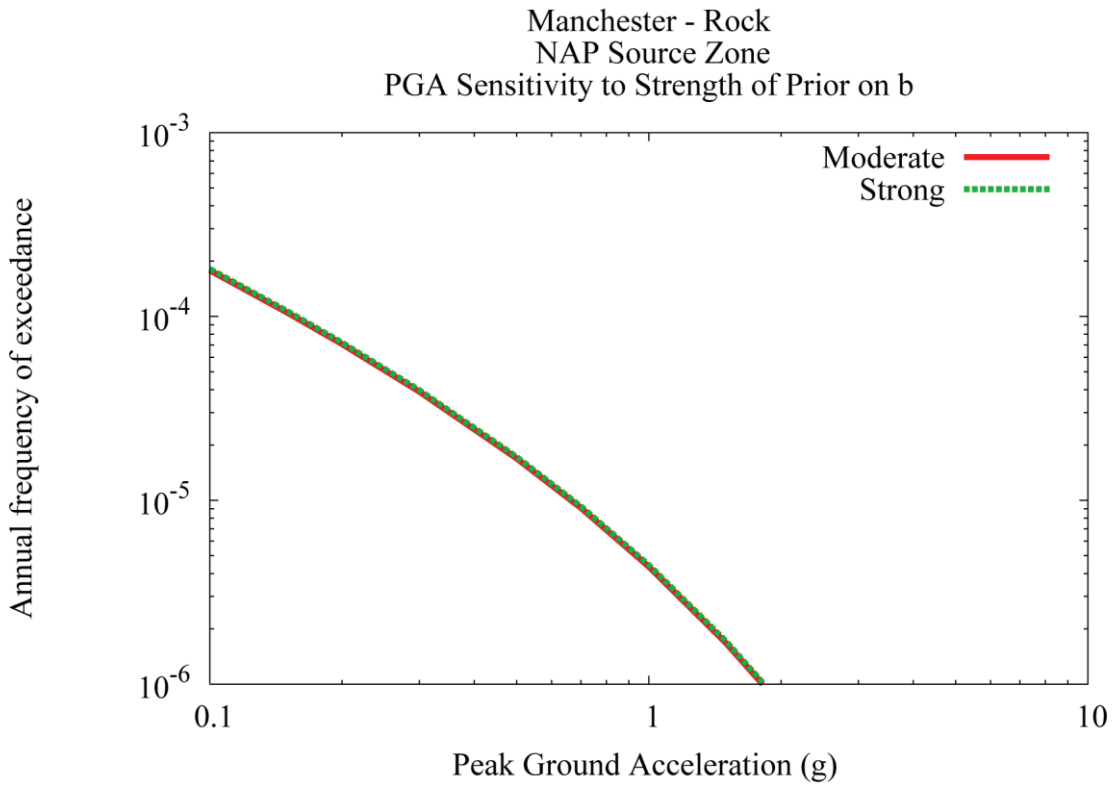

Figure 5.3.2-16
Sensitivity of seismic hazard at Manchester site to the strength of the prior on b

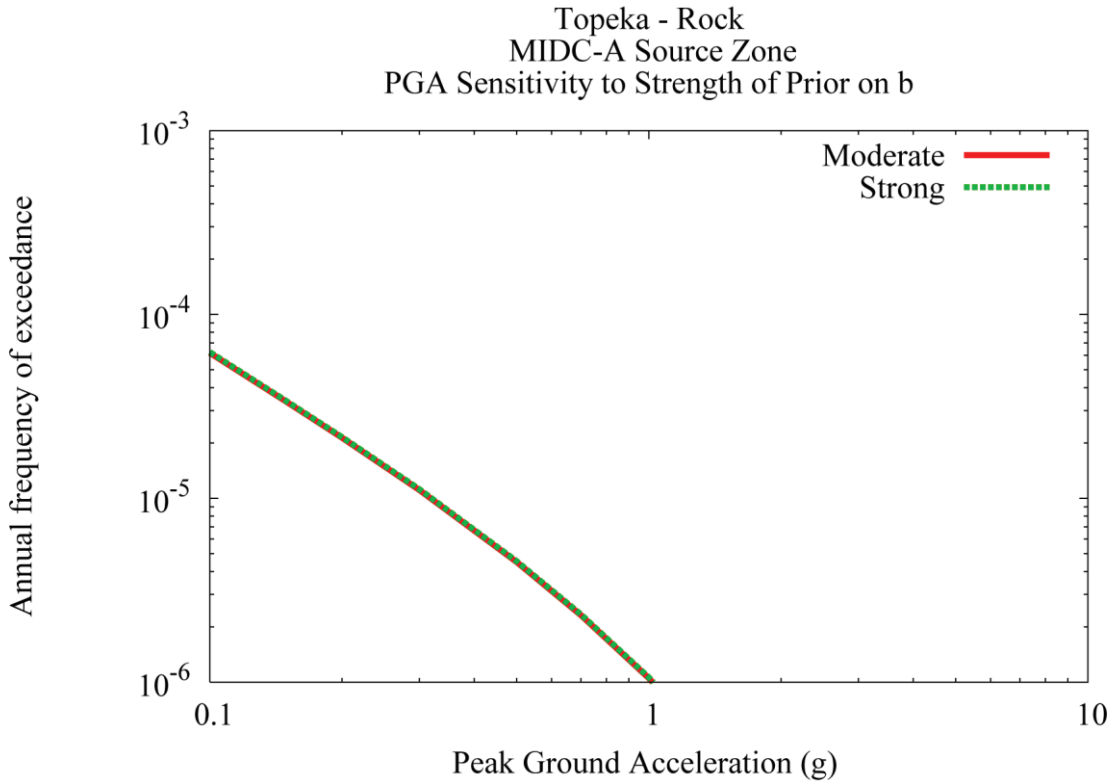

Figure 5.3.2-17
Sensitivity of seismic hazard at Topeka site to the strength of the prior on b

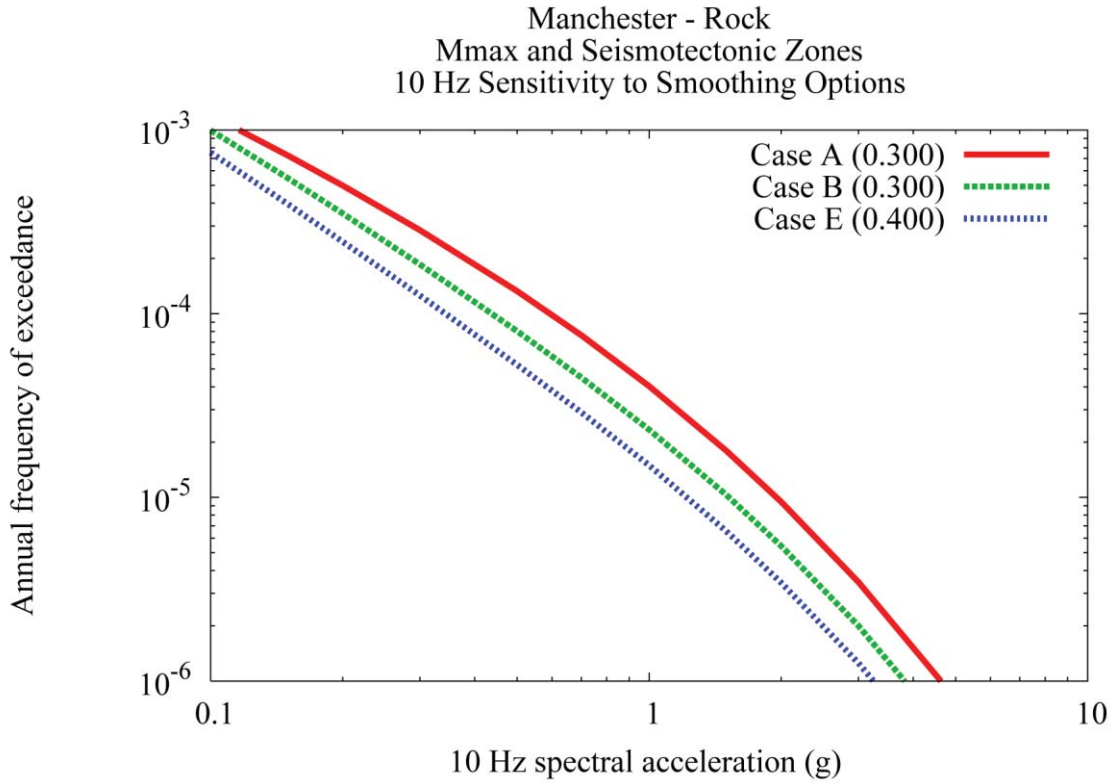

Figure 5.3.2-18
Sensitivity of seismic hazard at Manchester site to the choice of magnitude weights

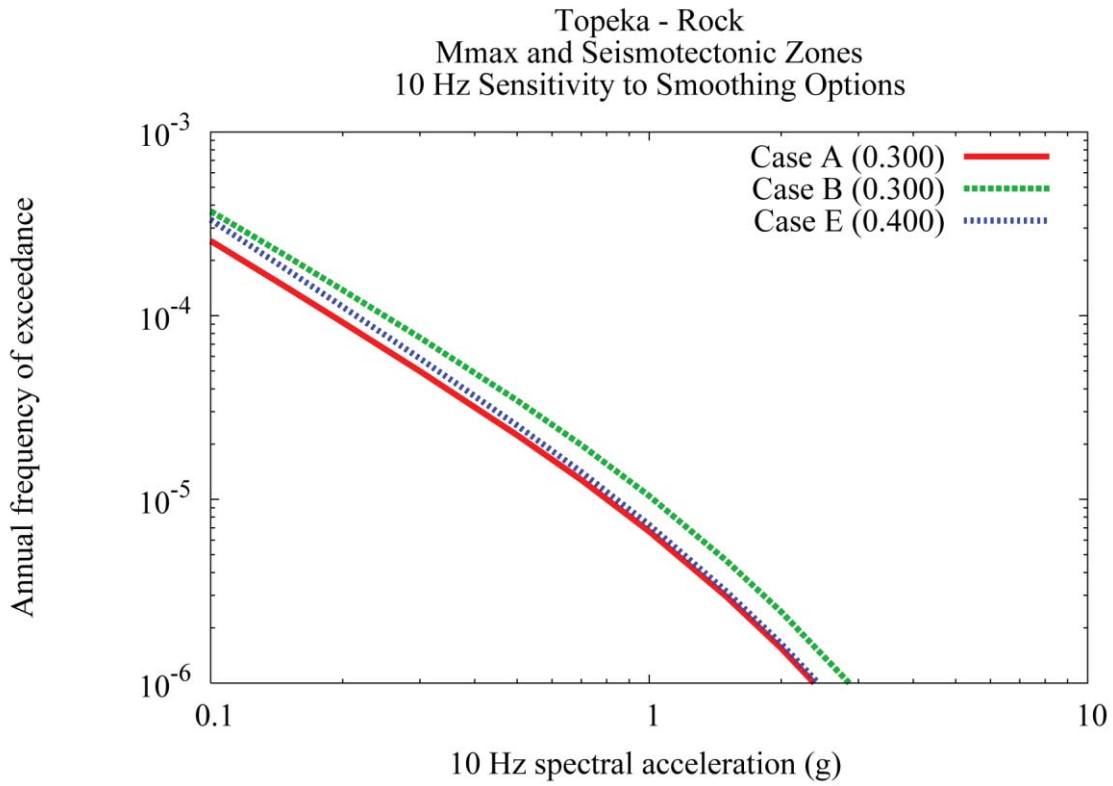

Figure 5.3.2-19
Sensitivity of seismic hazard at Topeka site to the choice of magnitude weights

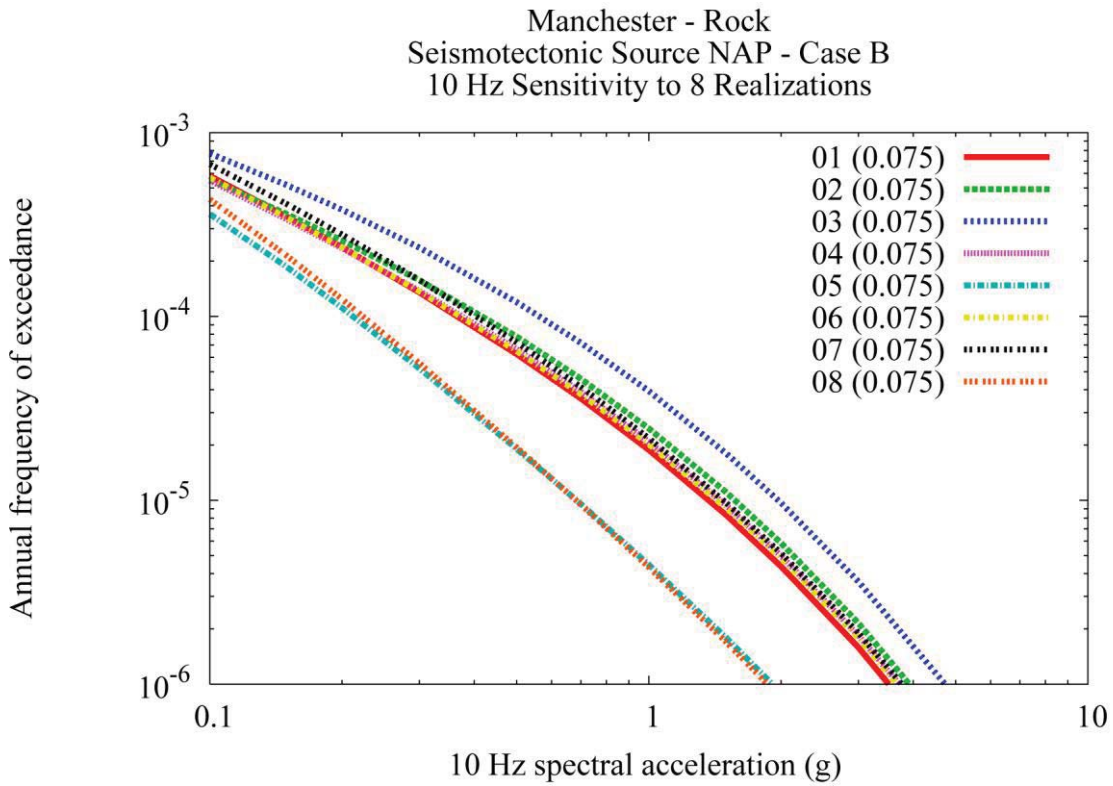

Figure 5.3.2-20
Sensitivity of seismic hazard from source NAP at Manchester site to the eight alternative recurrence maps for Case B magnitude weights

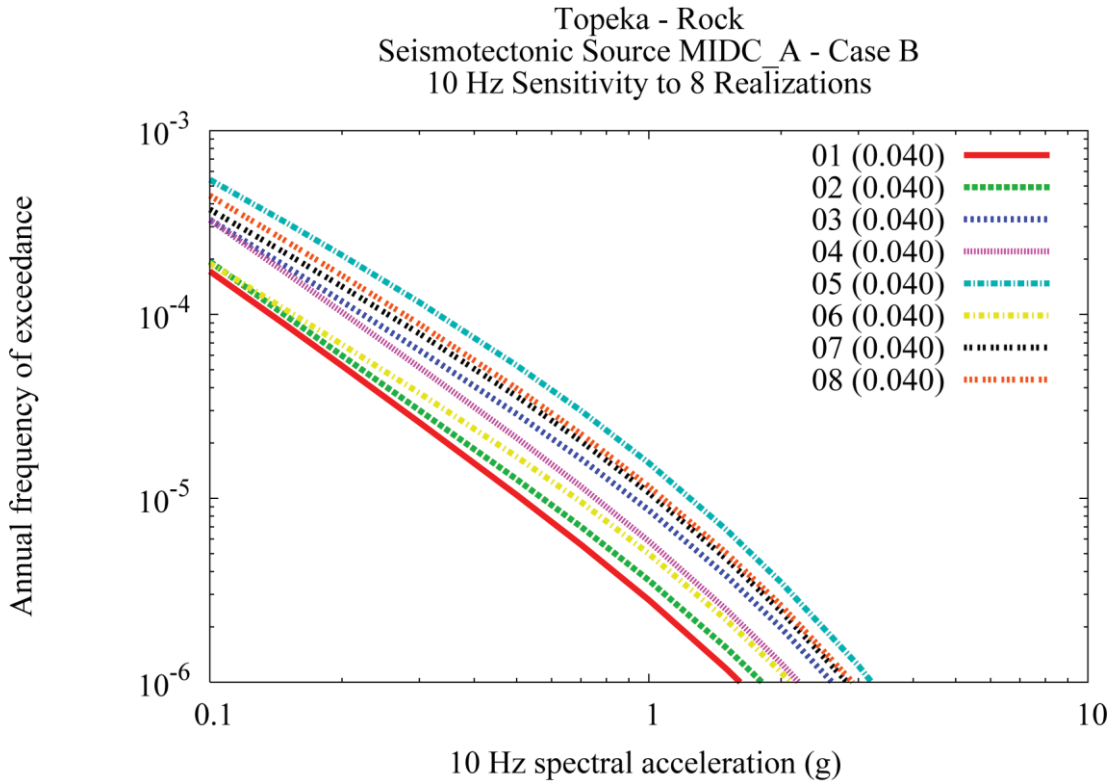

Figure 5.3.2-21
Sensitivity of seismic hazard from source MID-C–A at Topeka site to the eight alternative recurrence maps for Case B magnitude weights

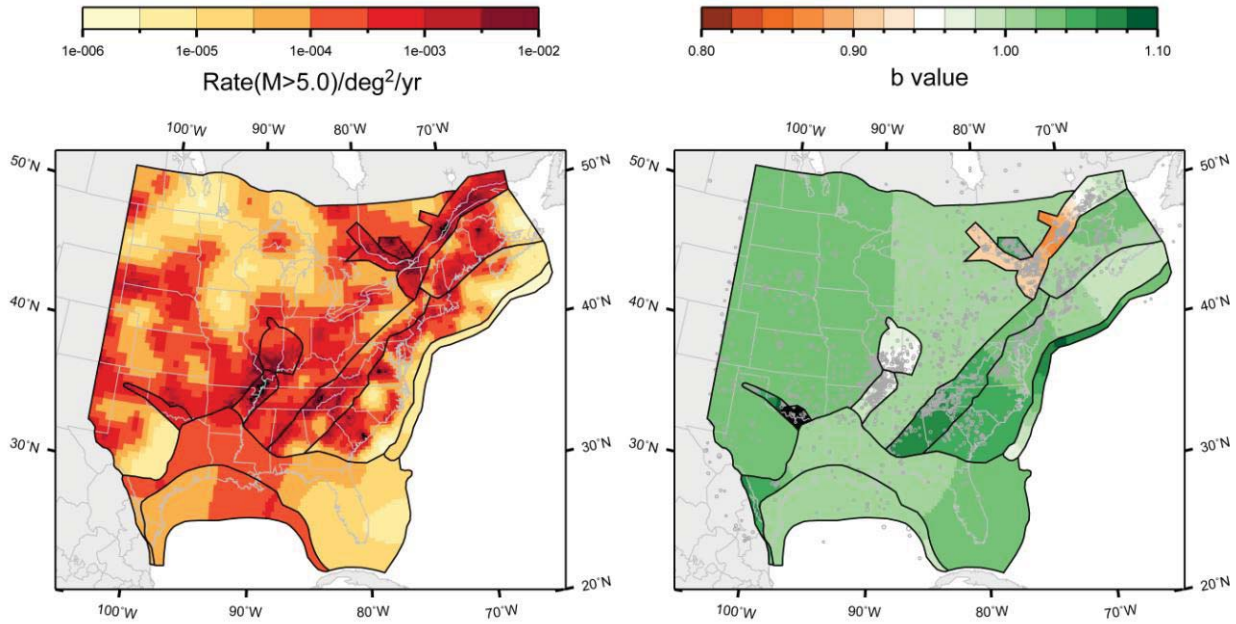

Figure 5.3.2-22
Mean recurrence-parameter map for the study region under the highest weighted source-zone configuration in the master logic tree. See Sections 6.3 and 7.5 for all mean maps.

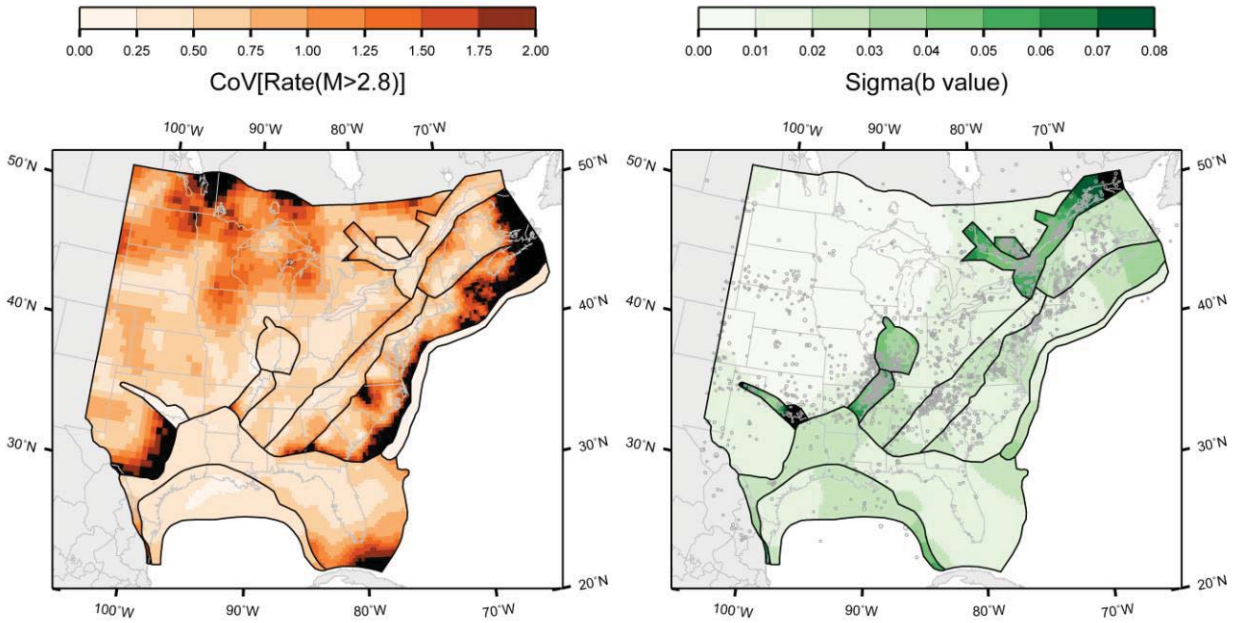

Figure 5.3.2-23
Map of the uncertainty in the estimated recurrence parameters, expressed as the coefficient of variation of the rate (left) and the standard deviation of the *b*-value (right) for the study region, under the highest weighted source-zone configuration in the master logic tree. See Appendix J for all maps of uncertainty.

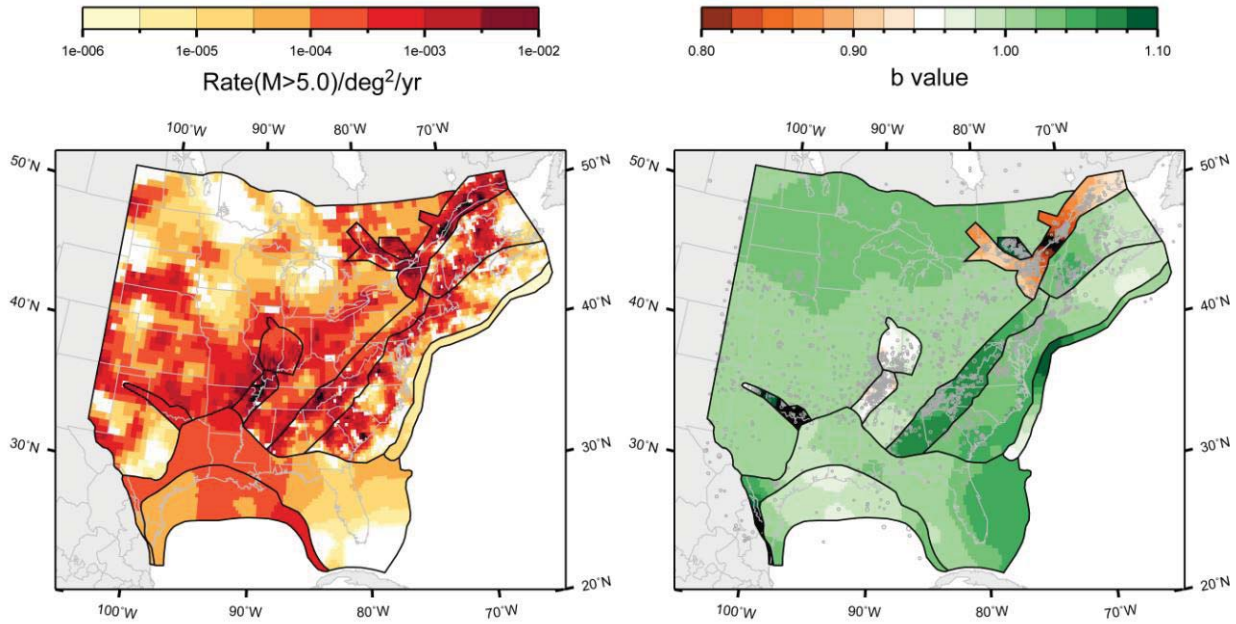

Figure 5.3.2-24
First of eight equally likely realizations of the recurrence-parameter map for the study region under the highest weighted source-zone configuration in the master logic tree. See Appendix J for maps of all realizations for all source-zone configurations.

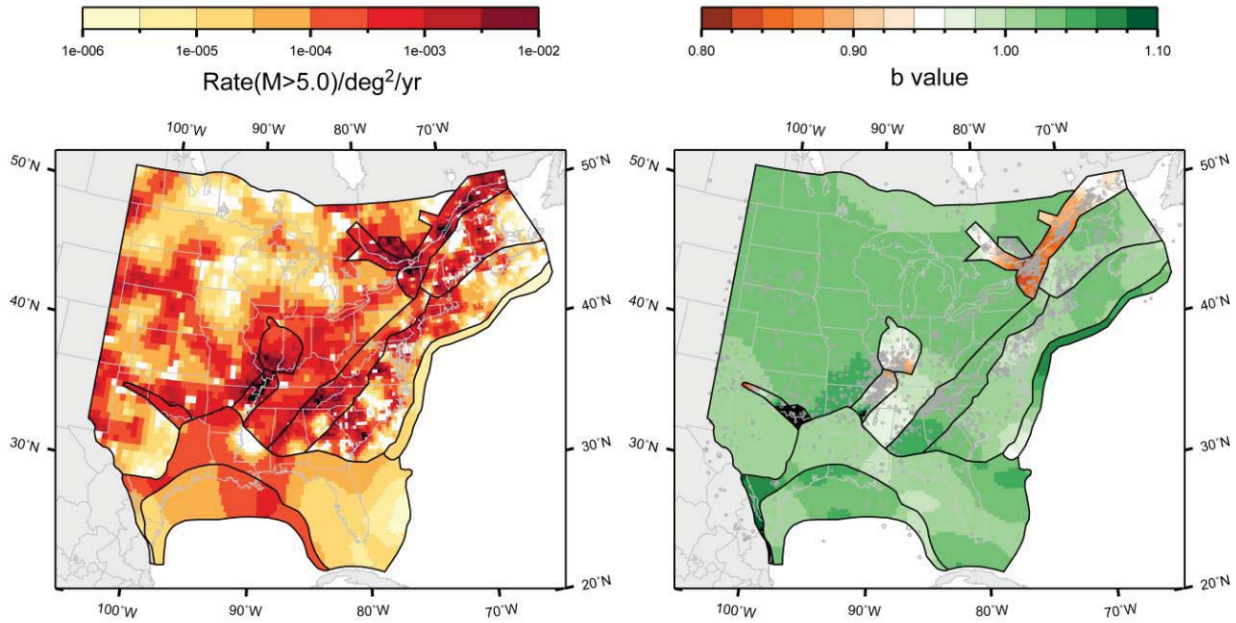

Figure 5.3.2-25
Eighth of eight equally likely realizations of the recurrence-parameter map for the study region under the highest weighted source-zone configuration in the master logic tree. See Appendix J for maps of all realizations for all source-zone configurations.

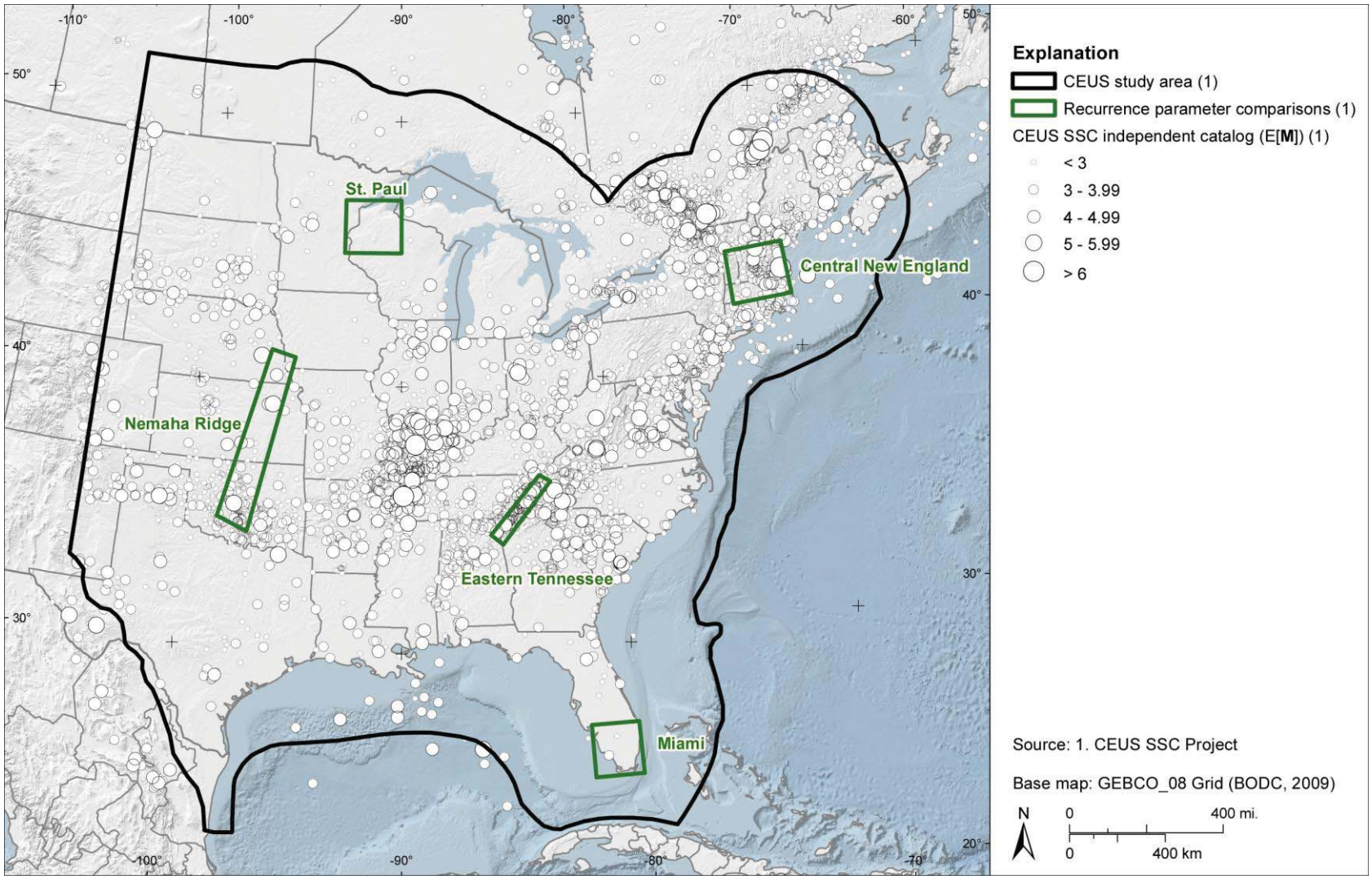

Figure 5.3.2-26
Map of geographic areas considered in the exploration of model results

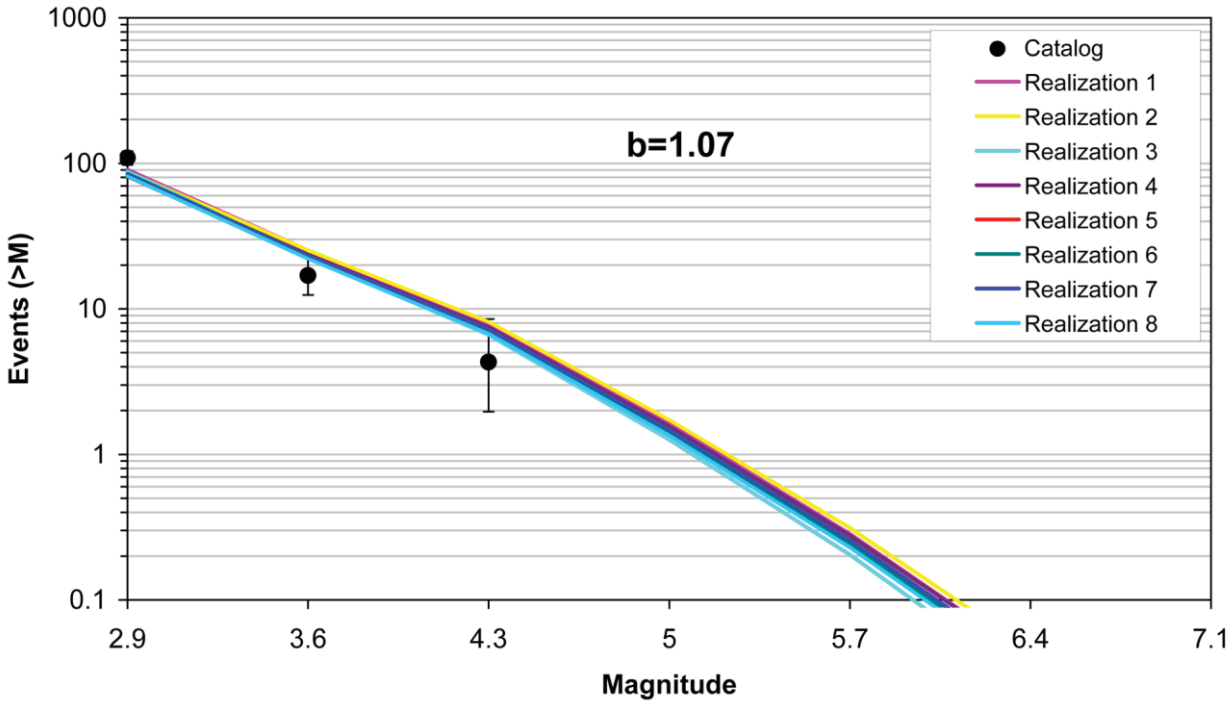

Figure 5.3.2-27
Comparison of model-predicted earthquake counts for the USGS Eastern Tennessee area using Case A magnitude weights. The error bars represent the 16%–84% uncertainty associated with the data, computed using the Weichert (1980) procedure.

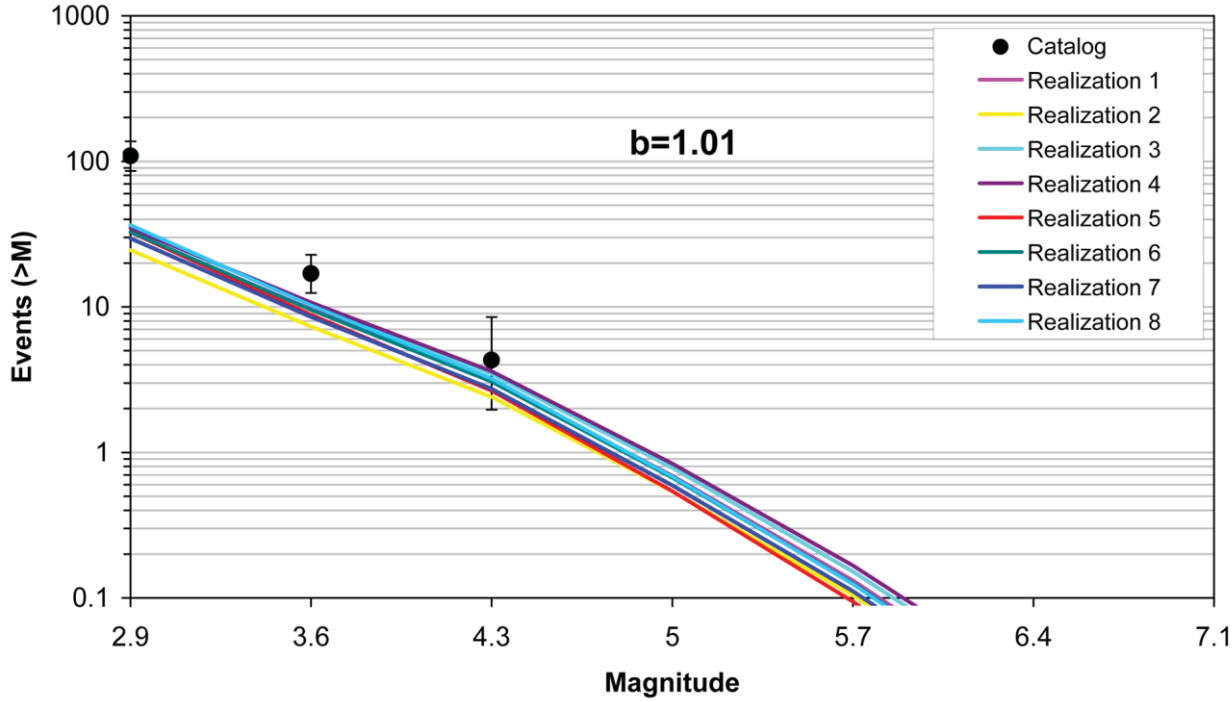

Figure 5.3.2-28
Comparison of model-predicted earthquake counts for the USGS Eastern Tennessee area using Case B magnitude weights

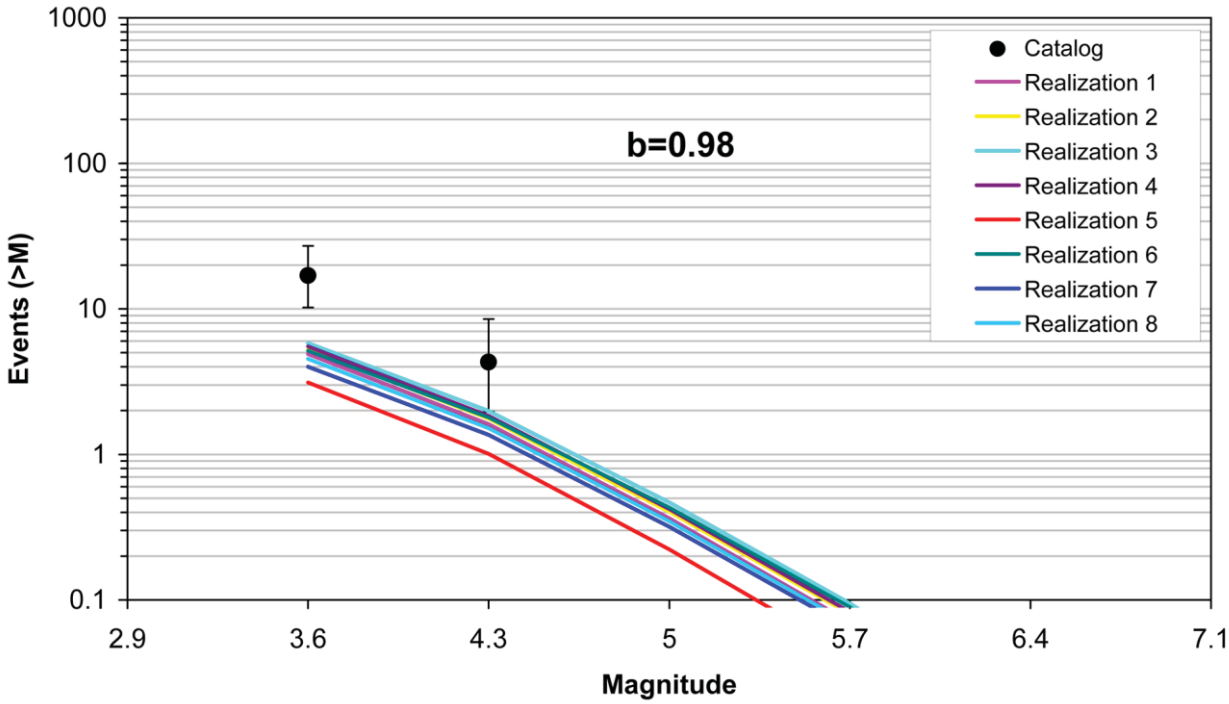

Figure 5.3.2-29
Comparison of model-predicted earthquake counts for the USGS Eastern Tennessee area using Case E magnitude weights

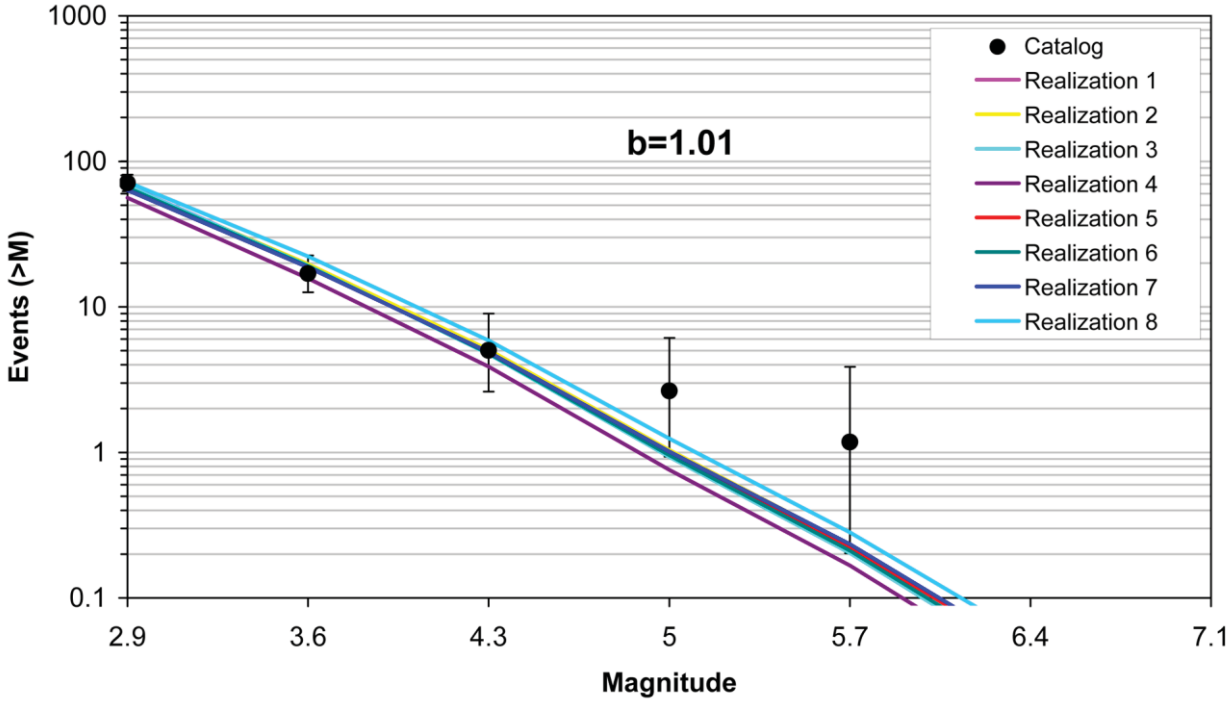

Figure 5.3.2-30
Comparison of model-predicted earthquake counts for the central New England area using Case A magnitude weights

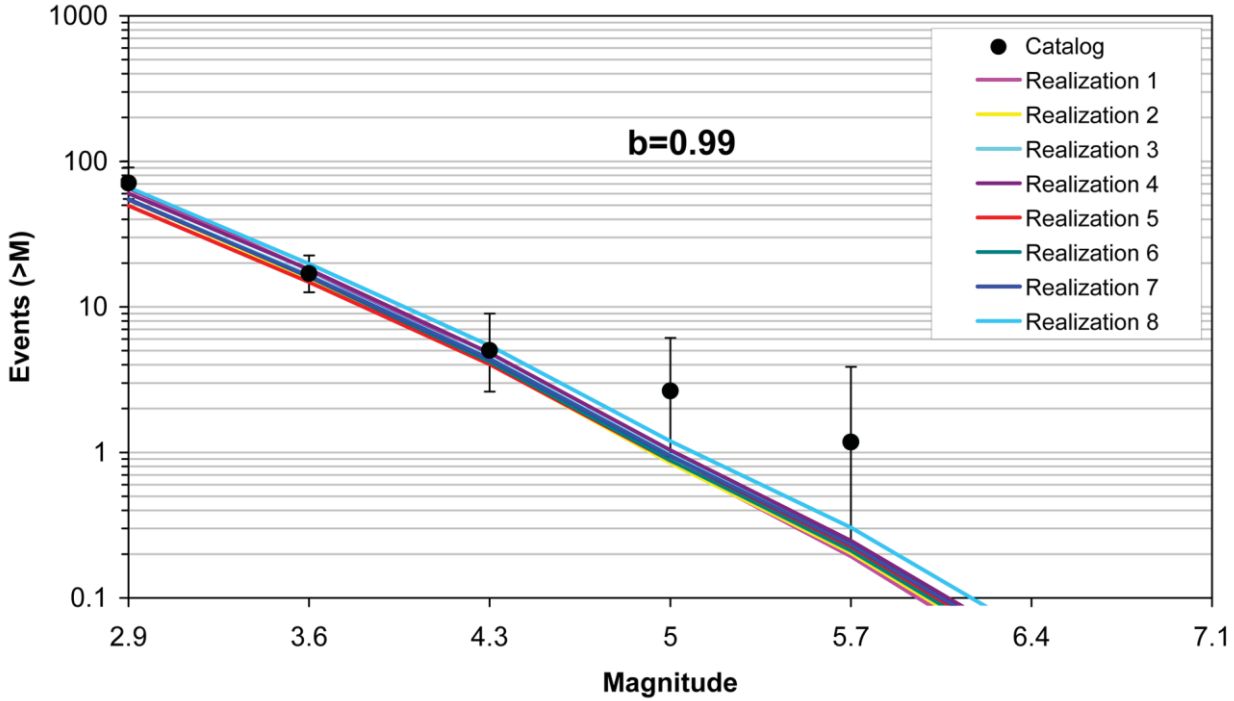

Figure 5.3.2-31
Comparison of model-predicted earthquake counts for the central New England area using Case B magnitude weights

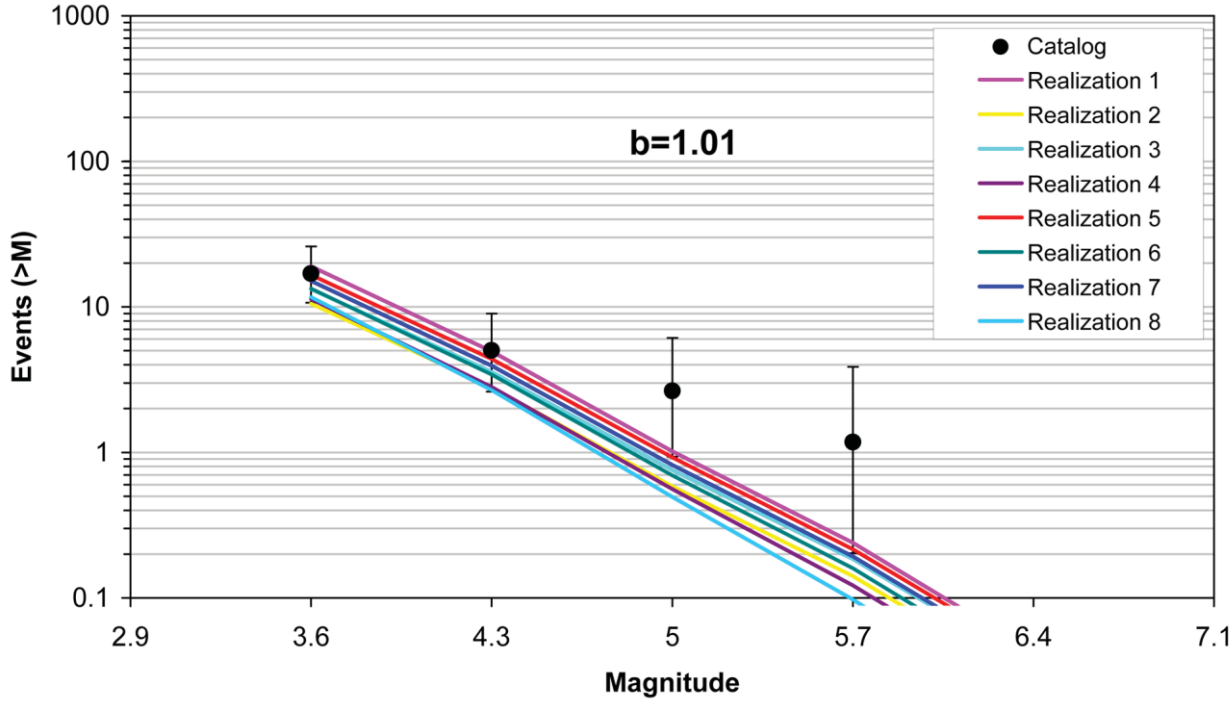

Figure 5.3.2-32
Comparison of model-predicted earthquake counts for the central New England area using Case E magnitude weights

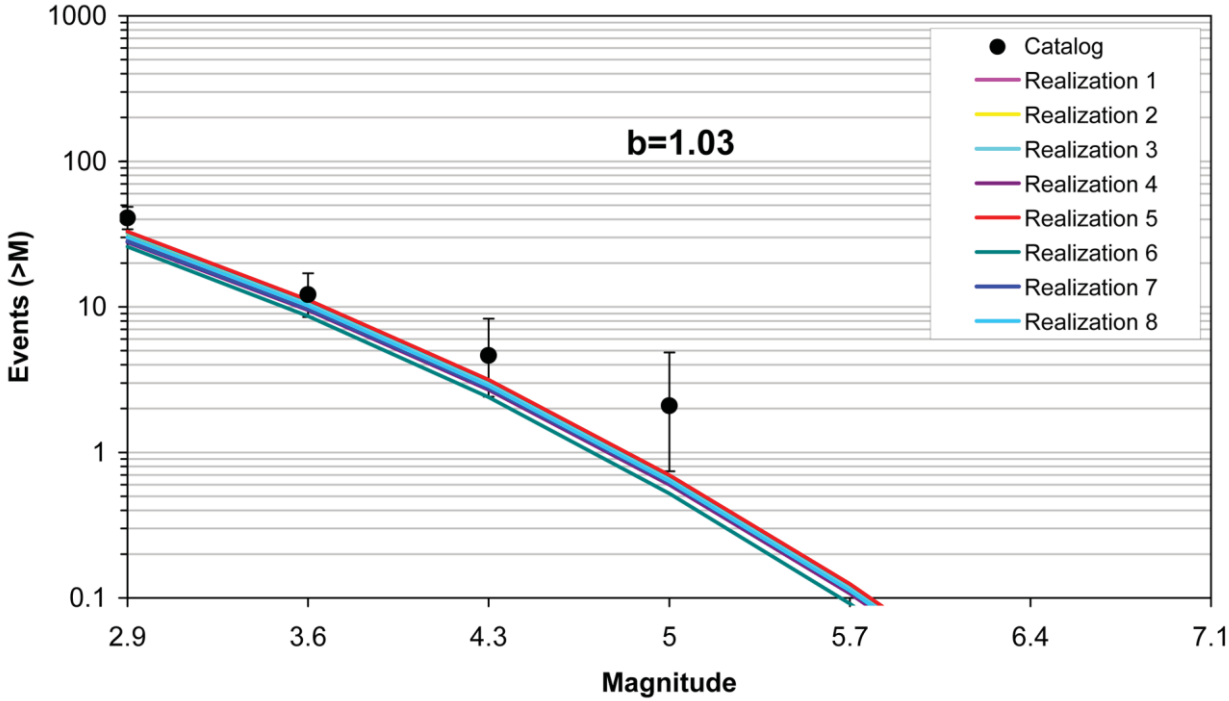

Figure 5.3.2-33
Comparison of model-predicted earthquake counts for the Nemaha Ridge area using Case A magnitude weights

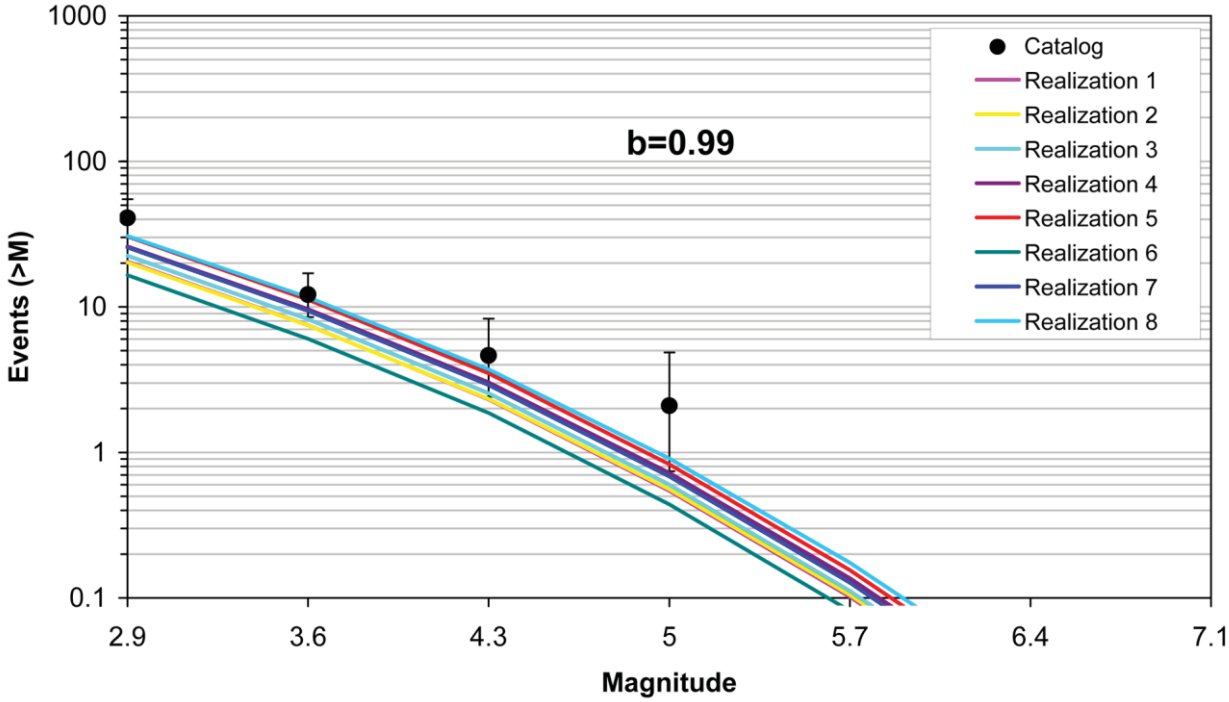

Figure 5.3.2-34
Comparison of model-predicted earthquake counts for the Nemaha Ridge area using Case B magnitude weights

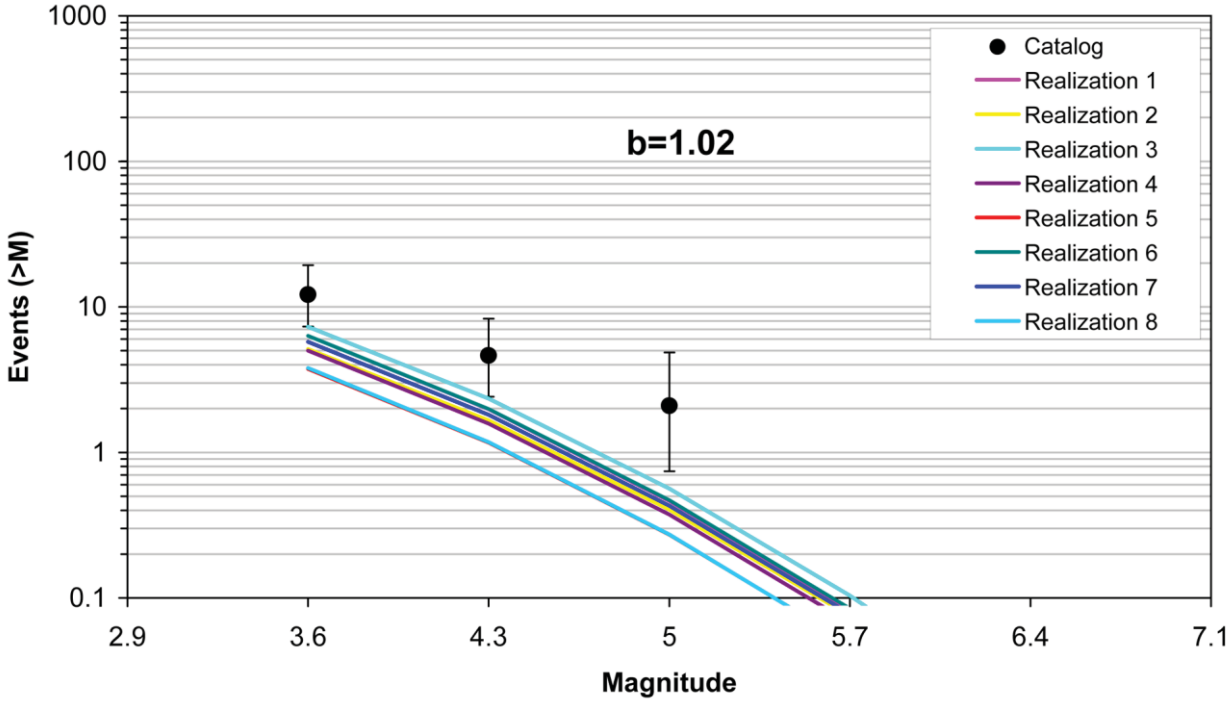

Figure 5.3.2-35
Comparison of model-predicted earthquake counts for the Nemaha Ridge area using Case E magnitude weights

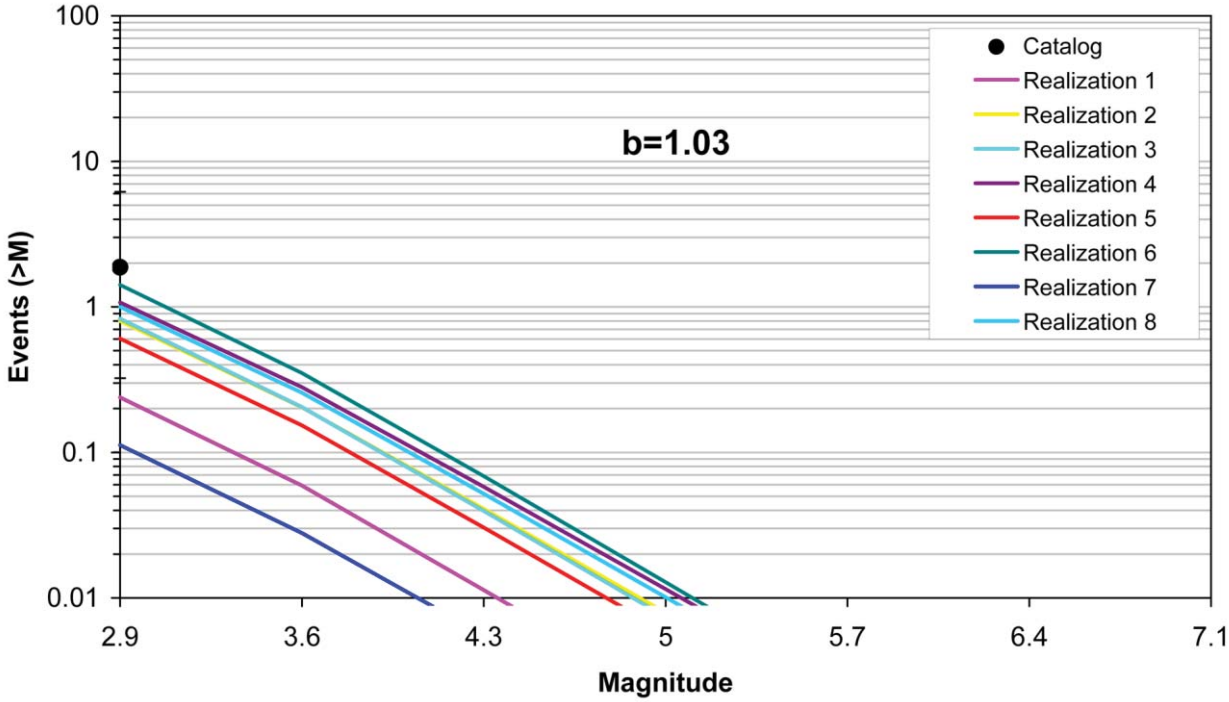

Figure 5.3.2-36
Comparison of model-predicted earthquake counts for the Miami, FL, area using Case A magnitude weights

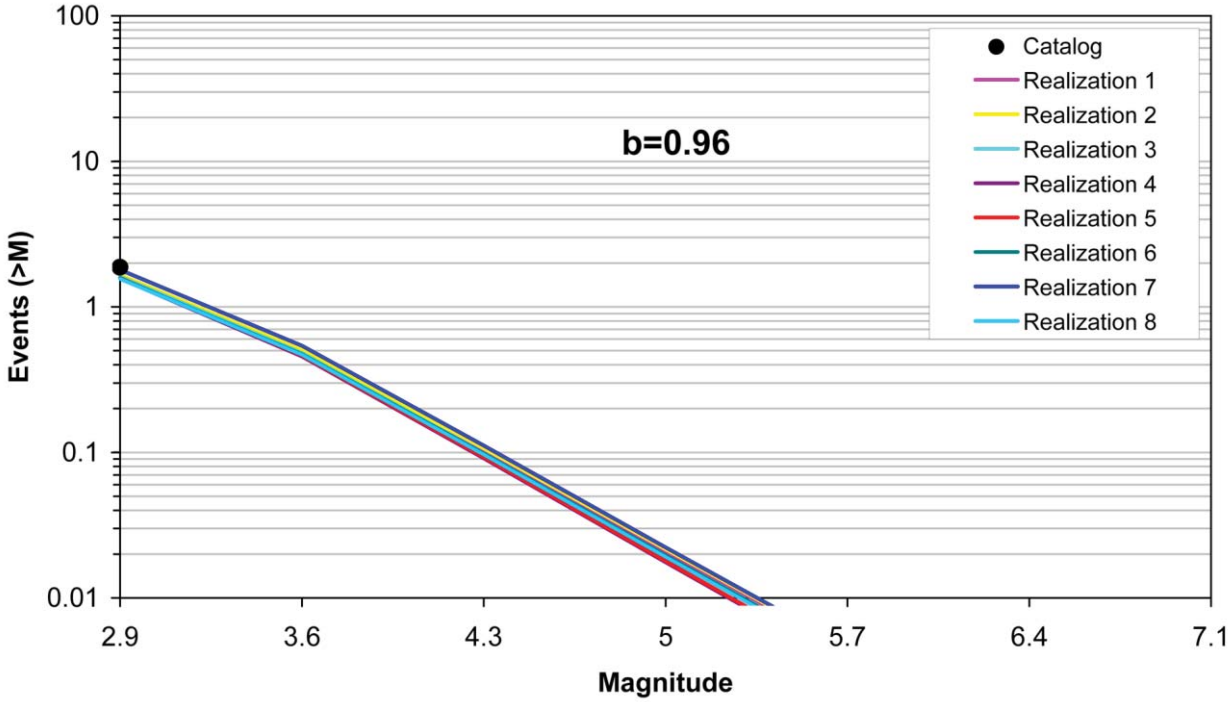

Figure 5.3.2-37
Comparison of model-predicted earthquake counts for the Miami, FL, area using Case B magnitude weights

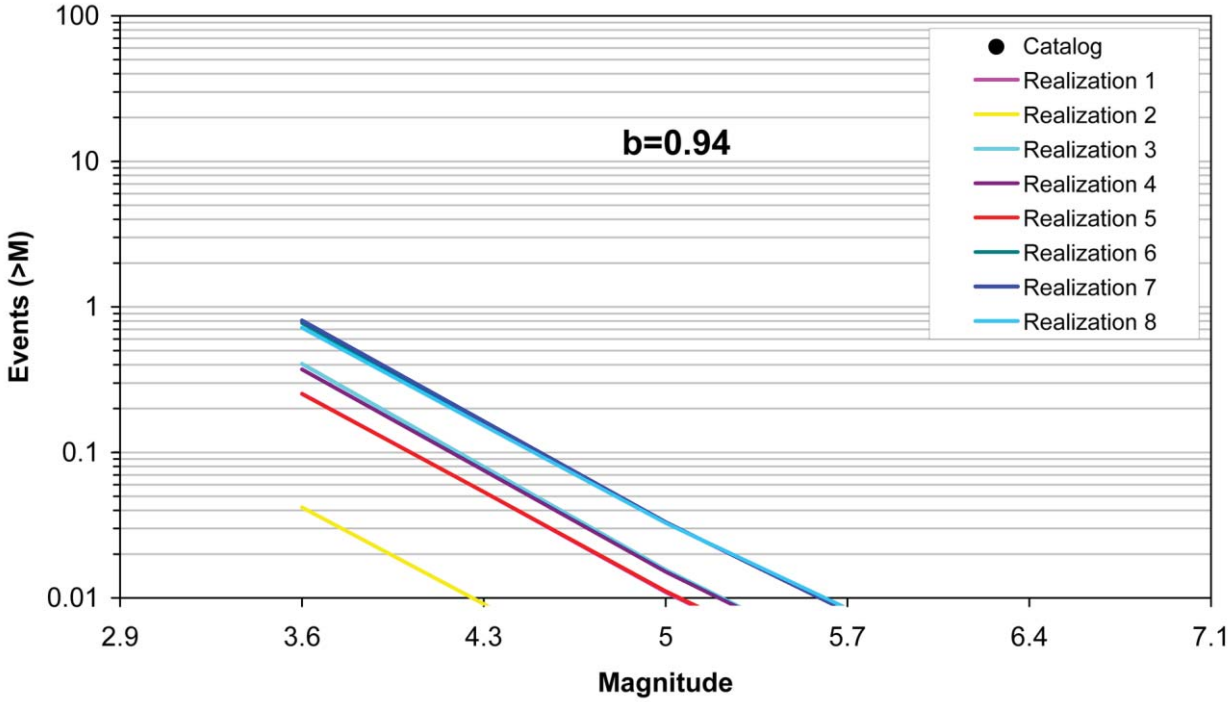

Figure 5.3.2-38
Comparison of model-predicted earthquake counts for the Miami, FL, area using Case E magnitude weights

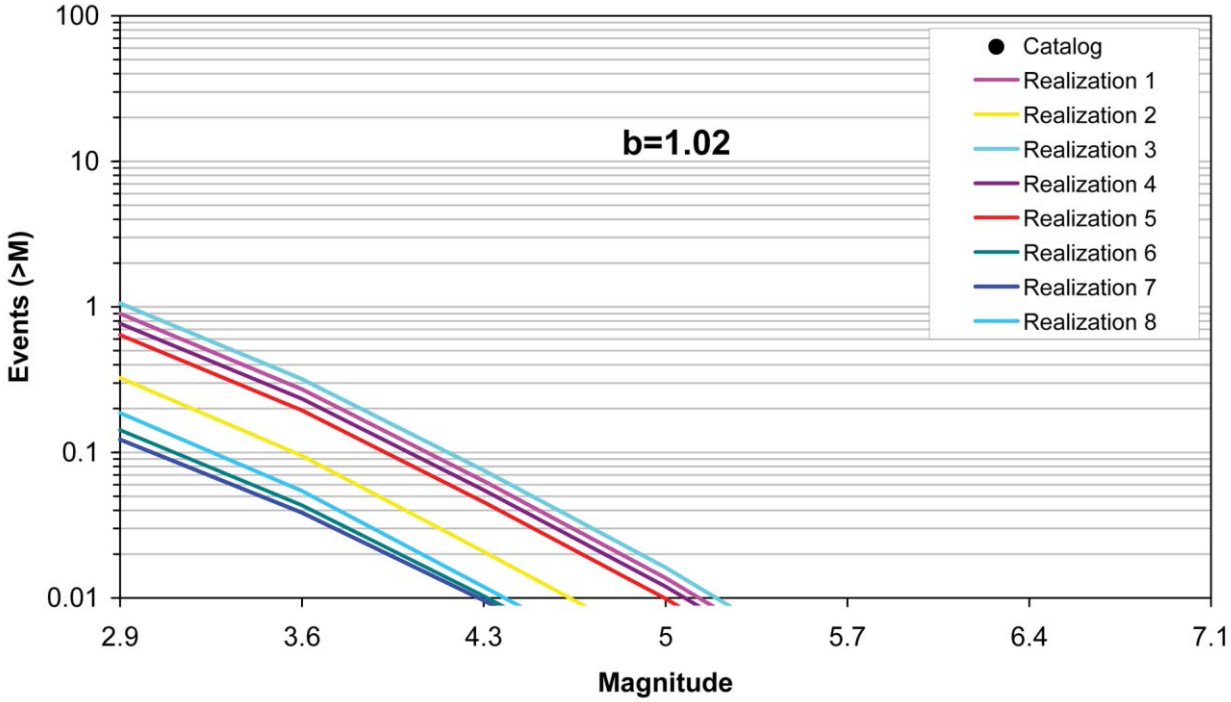

Figure 5.3.2-39
Comparison of model-predicted earthquake counts for the St. Paul, MN, area using Case A magnitude weights

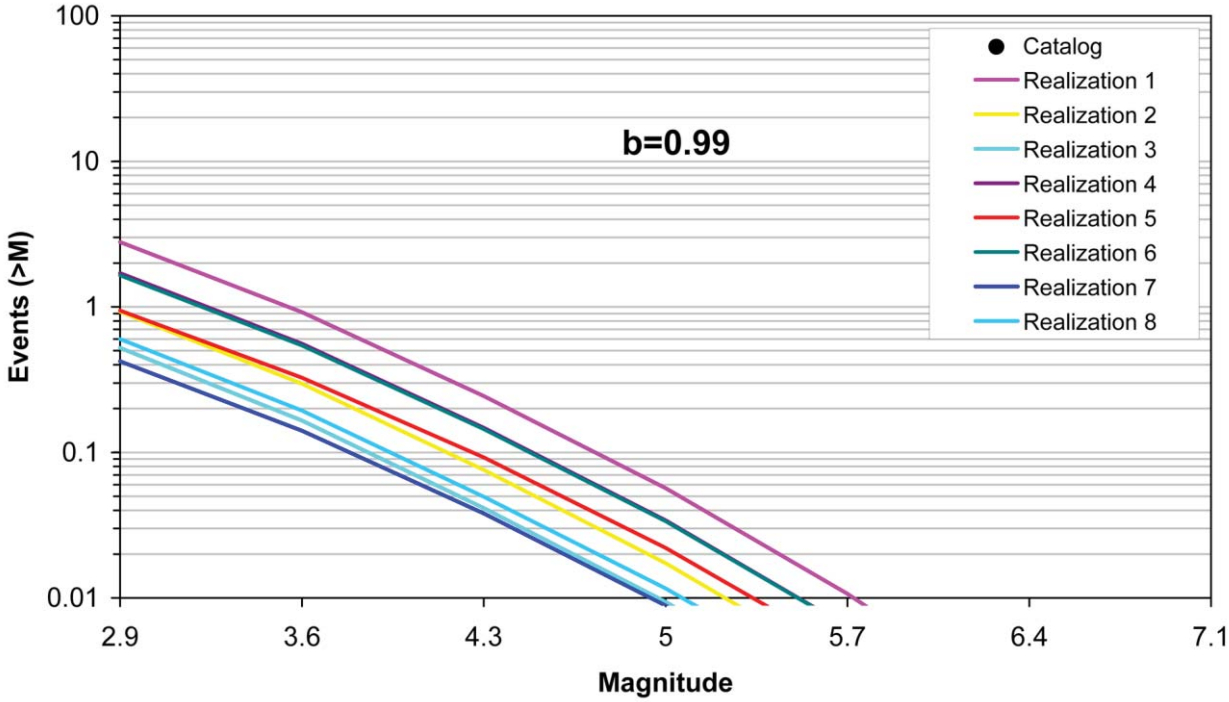

Figure 5.3.2-40
Comparison of model-predicted earthquake counts for the St. Paul, MN, area using Case B magnitude weights

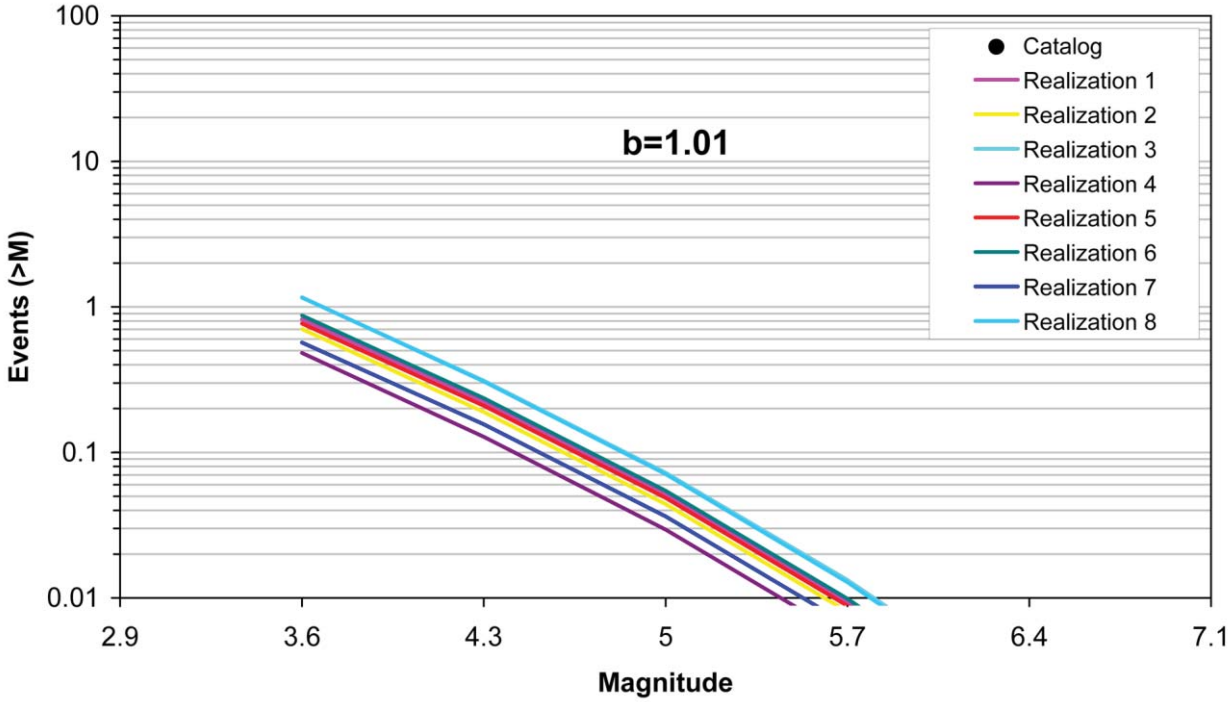

Figure 5.3.2-41
Comparison of model-predicted earthquake counts for the St. Paul, MN, area using Case E magnitude weights

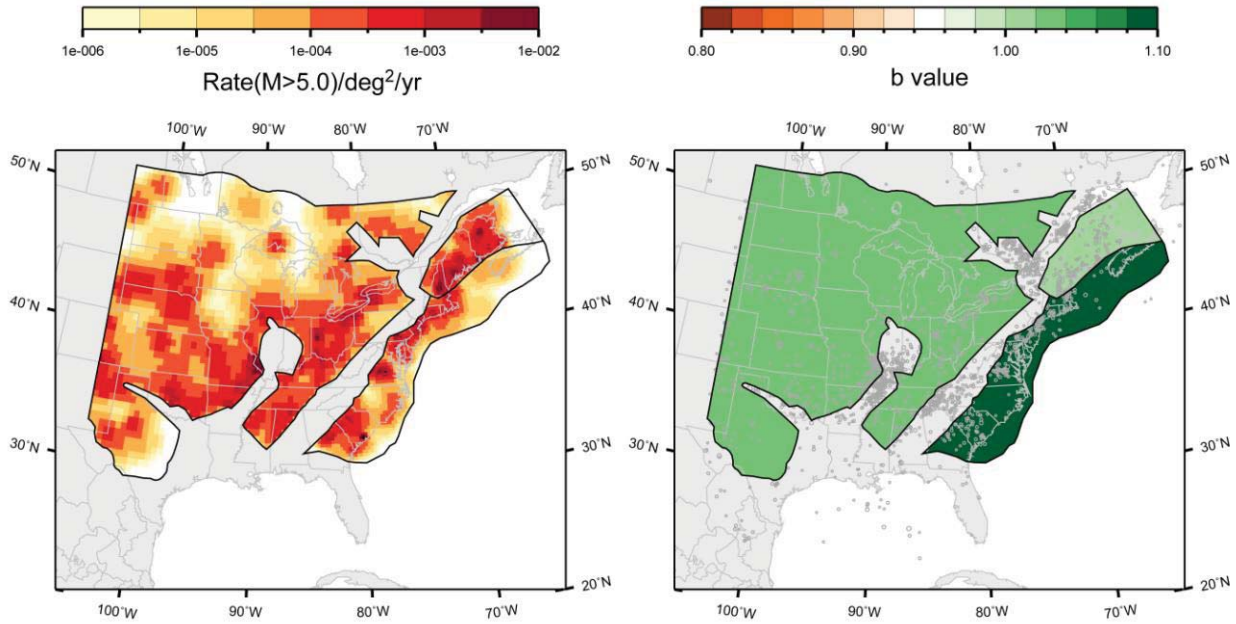

Figure 5.3.2-42
Recurrence parameters for the ECC-AM, MID-C–A, and NAP seismotectonic source zones and Case A magnitude weights computed using an objective adaptive kernel approach

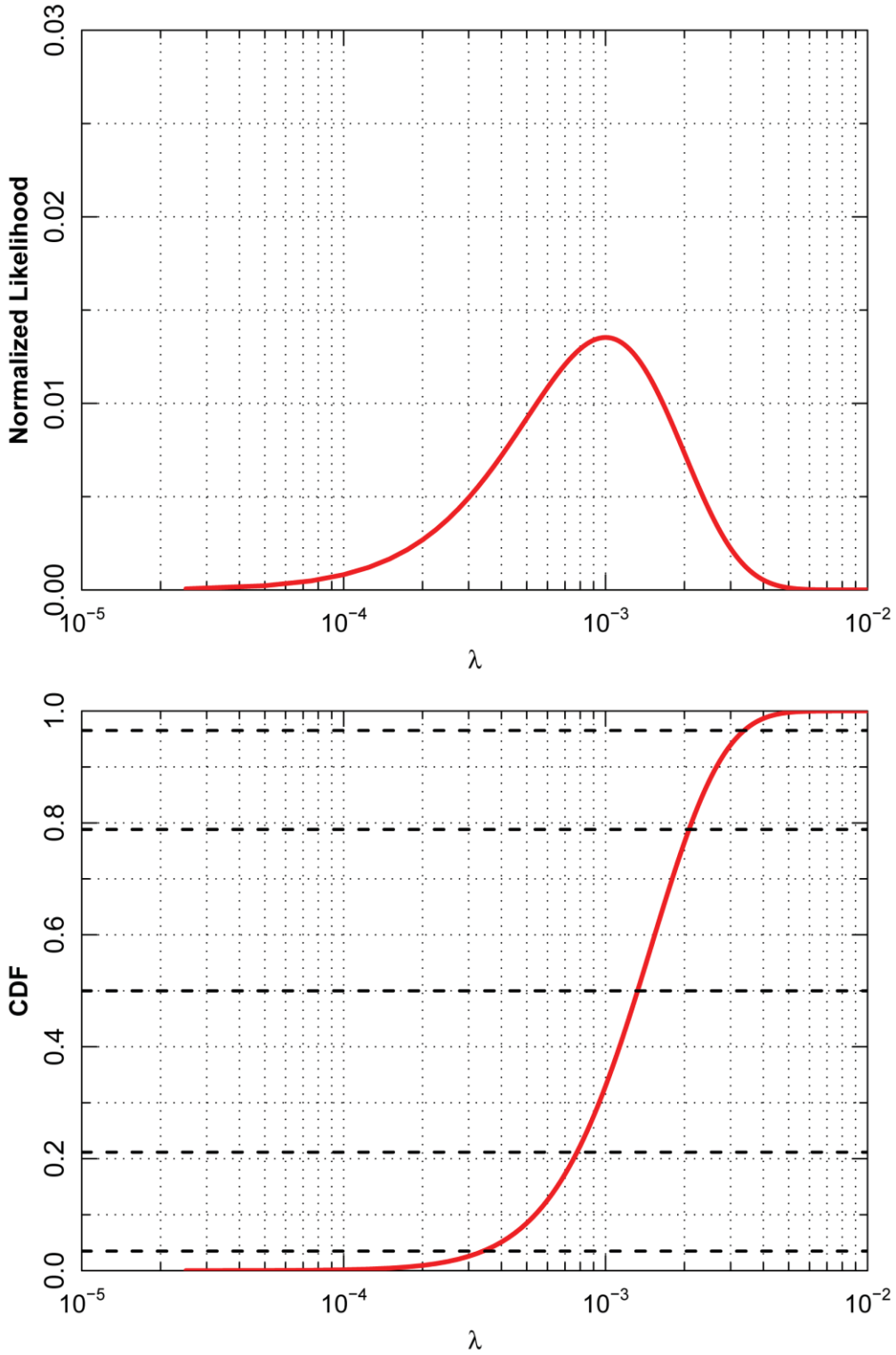

Figure 5.3.3-1
Likelihood distribution for rate parameter λ derived using Equation 5.3.3-1 for N = 2 and T = 2,000 years. Top: normalized probability density function for λ. Bottom: resulting cumulative distribution function. Dashed lines show the cumulative probability levels for the Miller and Rice (1983) discrete approximation of a continuous probability distribution.

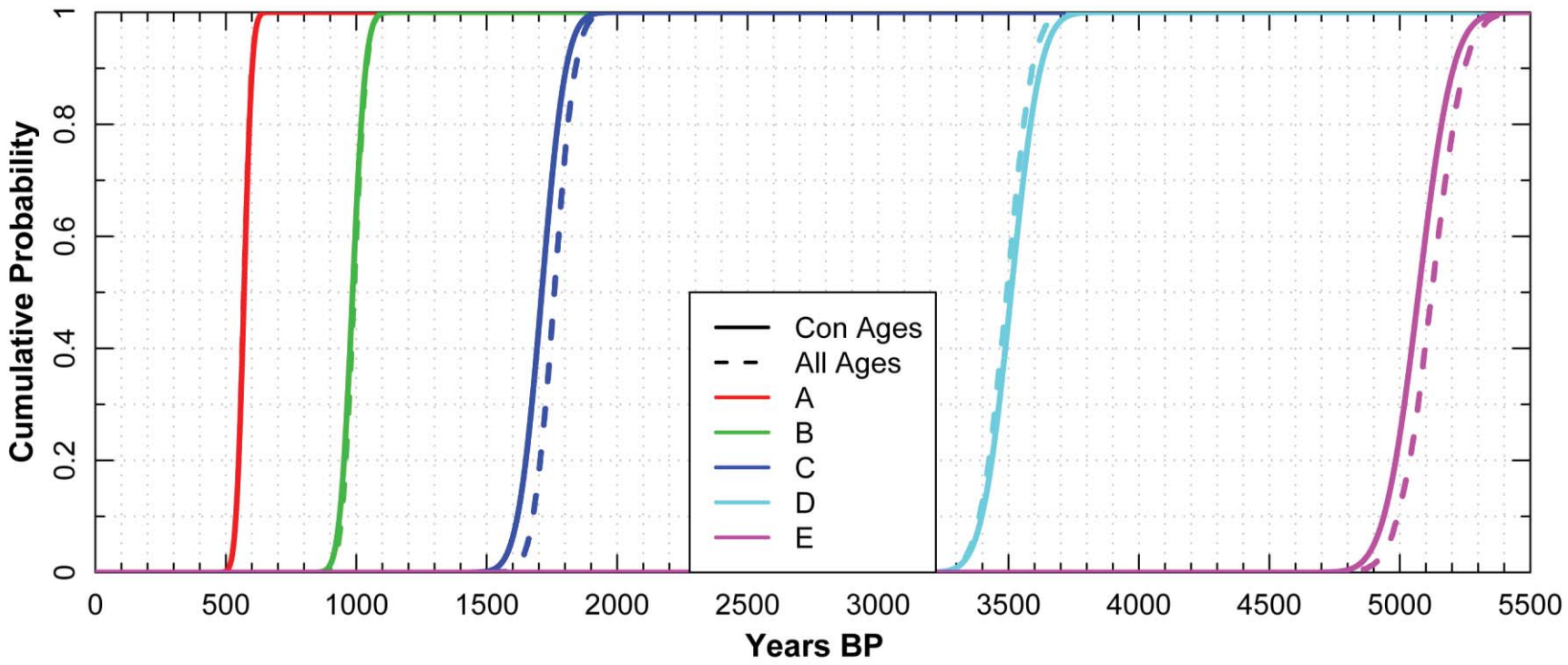

Figure 5.3.3-2
Uncertainty distributions for the age of Charleston RLMEs

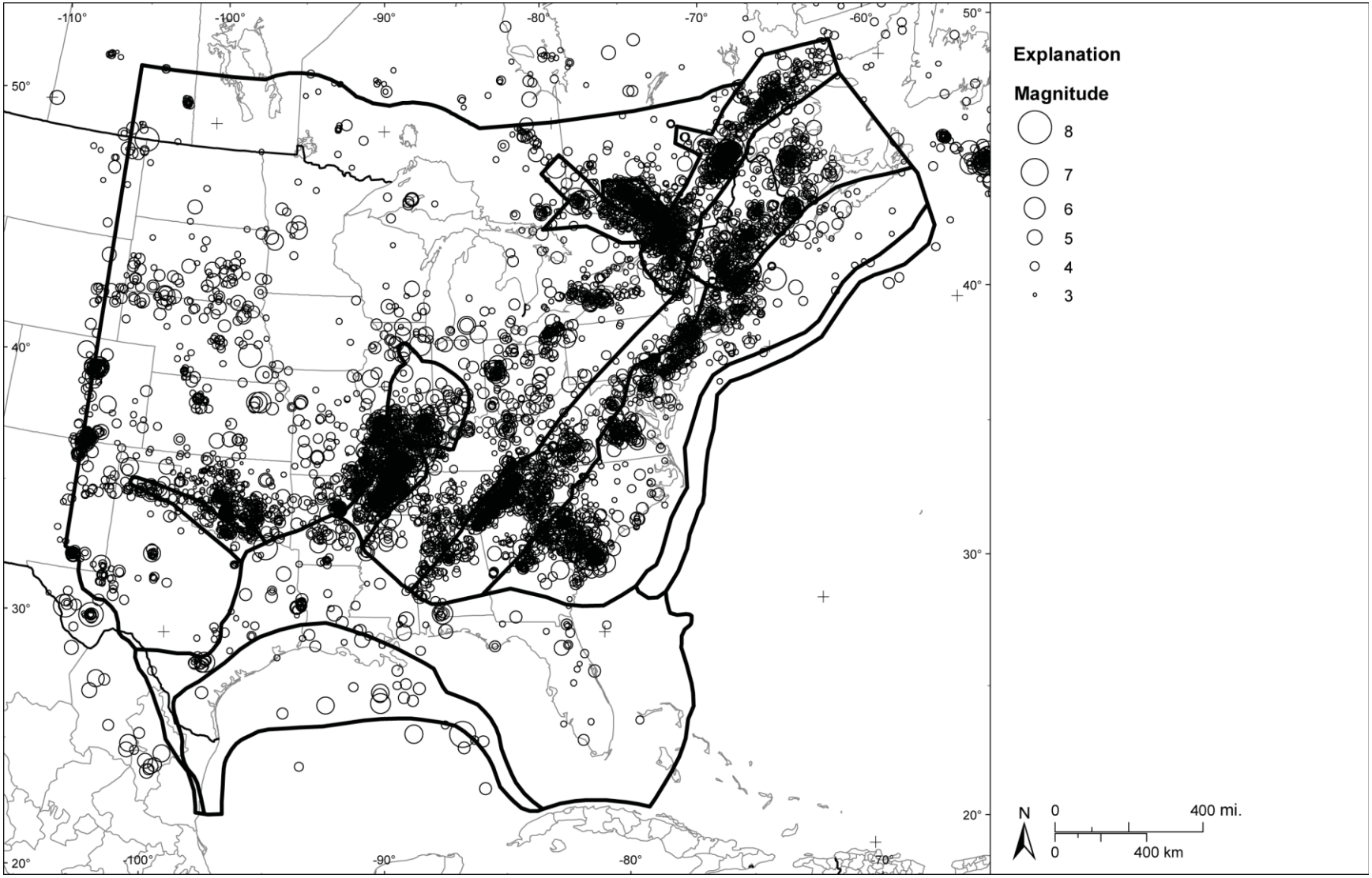

Figure 5.4.4-1
Spatial distribution of earthquakes in the CEUS SSC Project catalog. Solid lines indicate the boundaries of the seismotectonic source zones (narrow interpretation).

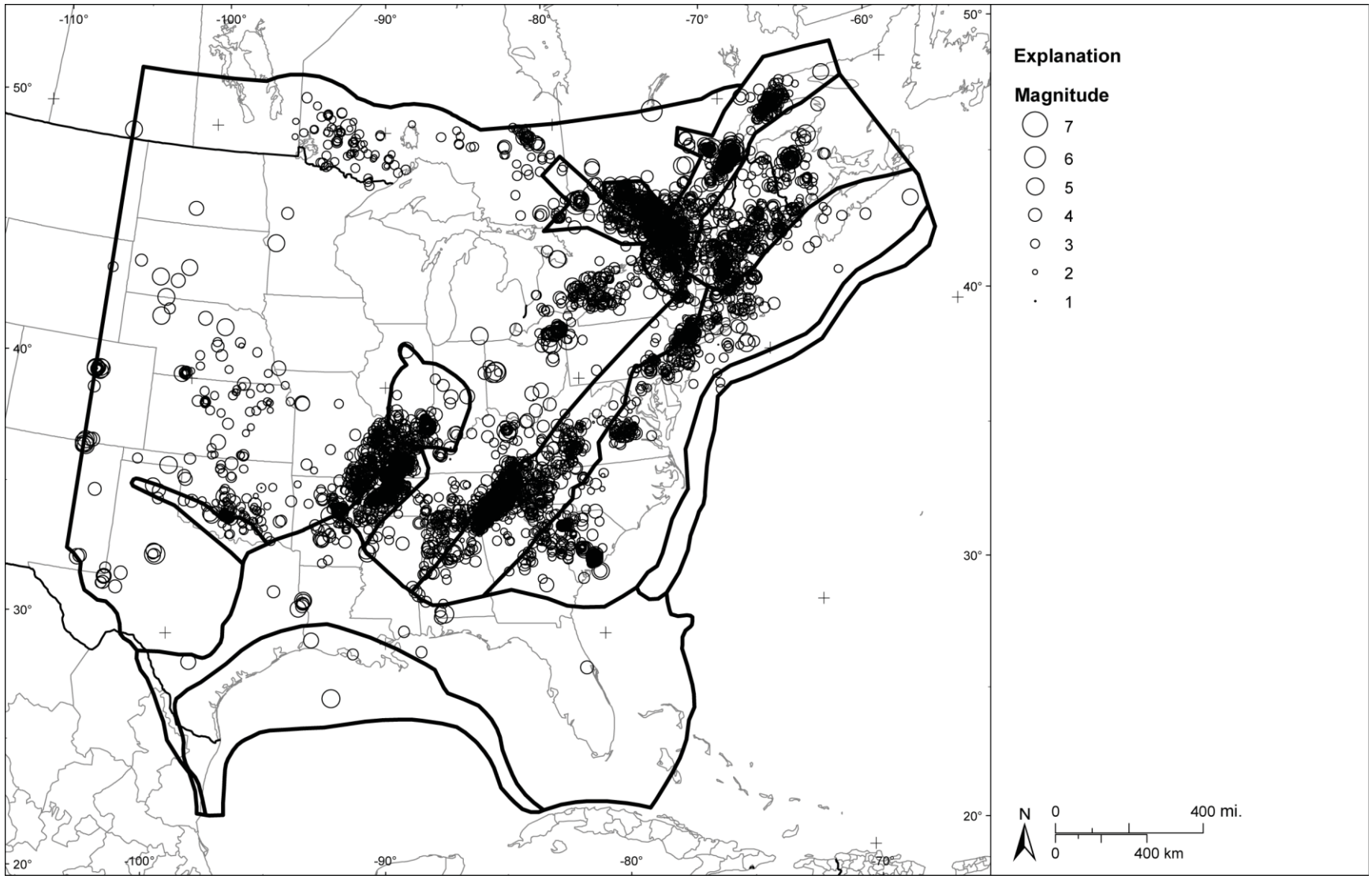

Figure 5.4.4-2
Spatial distribution of earthquakes in the CEUS SSC Project catalog with good quality depth determinations used for assessing crustal thickness. Solid lines indicate the boundaries of the seismotectonic source zones (narrow interpretation).

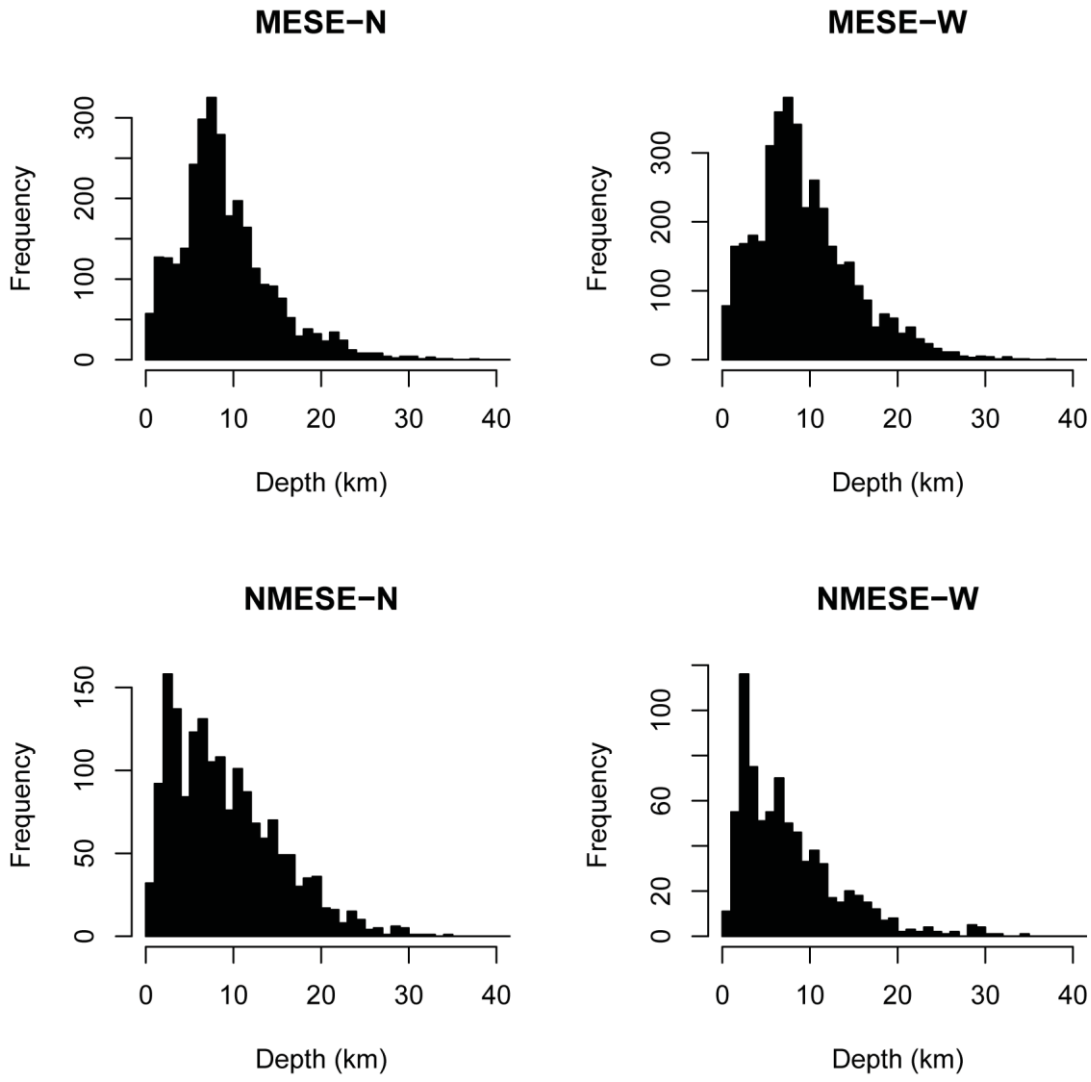

Figure 5.4.4-3
Distribution of better-quality focal depths in Mmax source zones

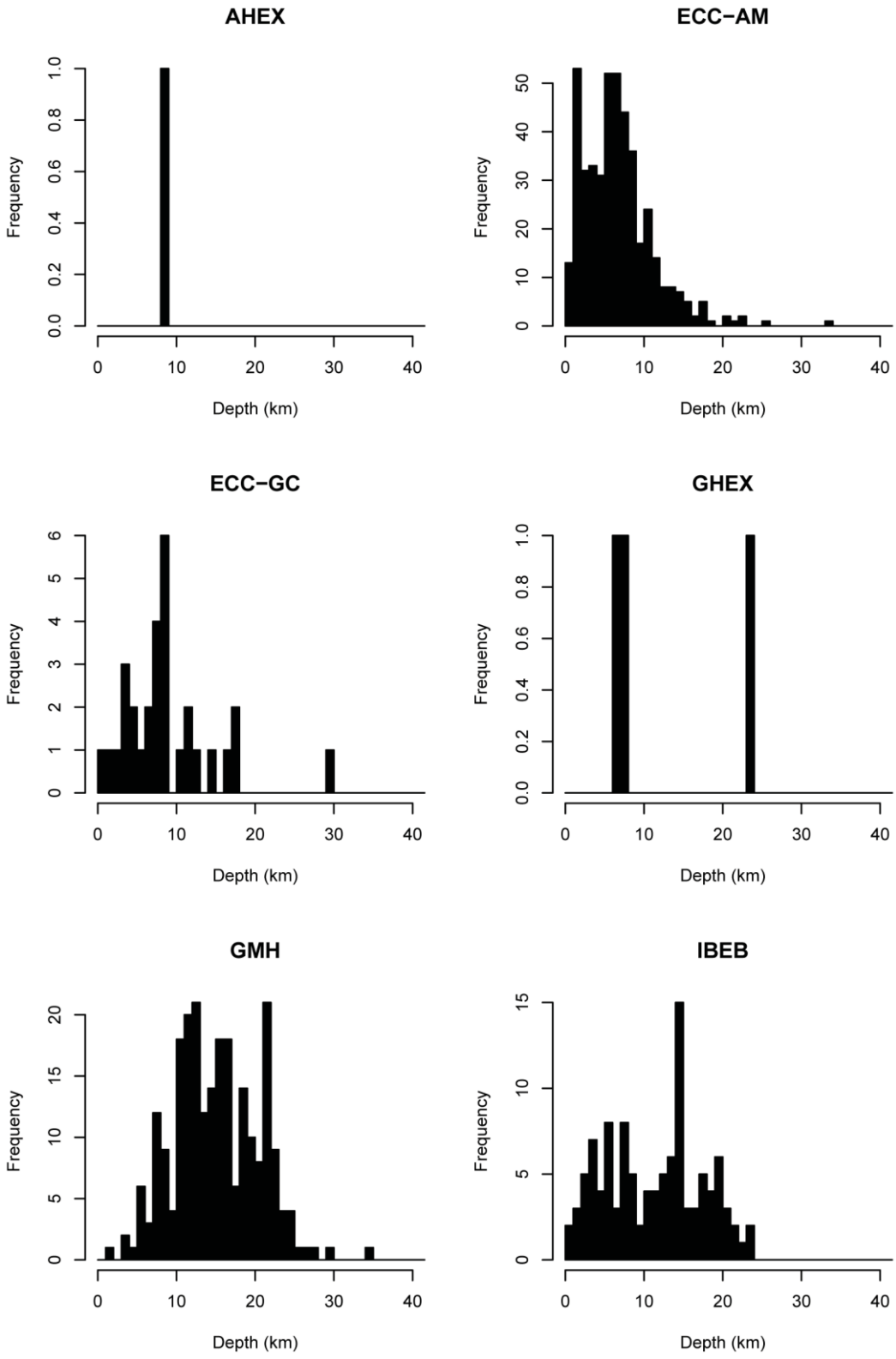

Figure 5.4.4-4 (1 of 3)
Distribution of better-quality focal depths in seismotectonic source zones

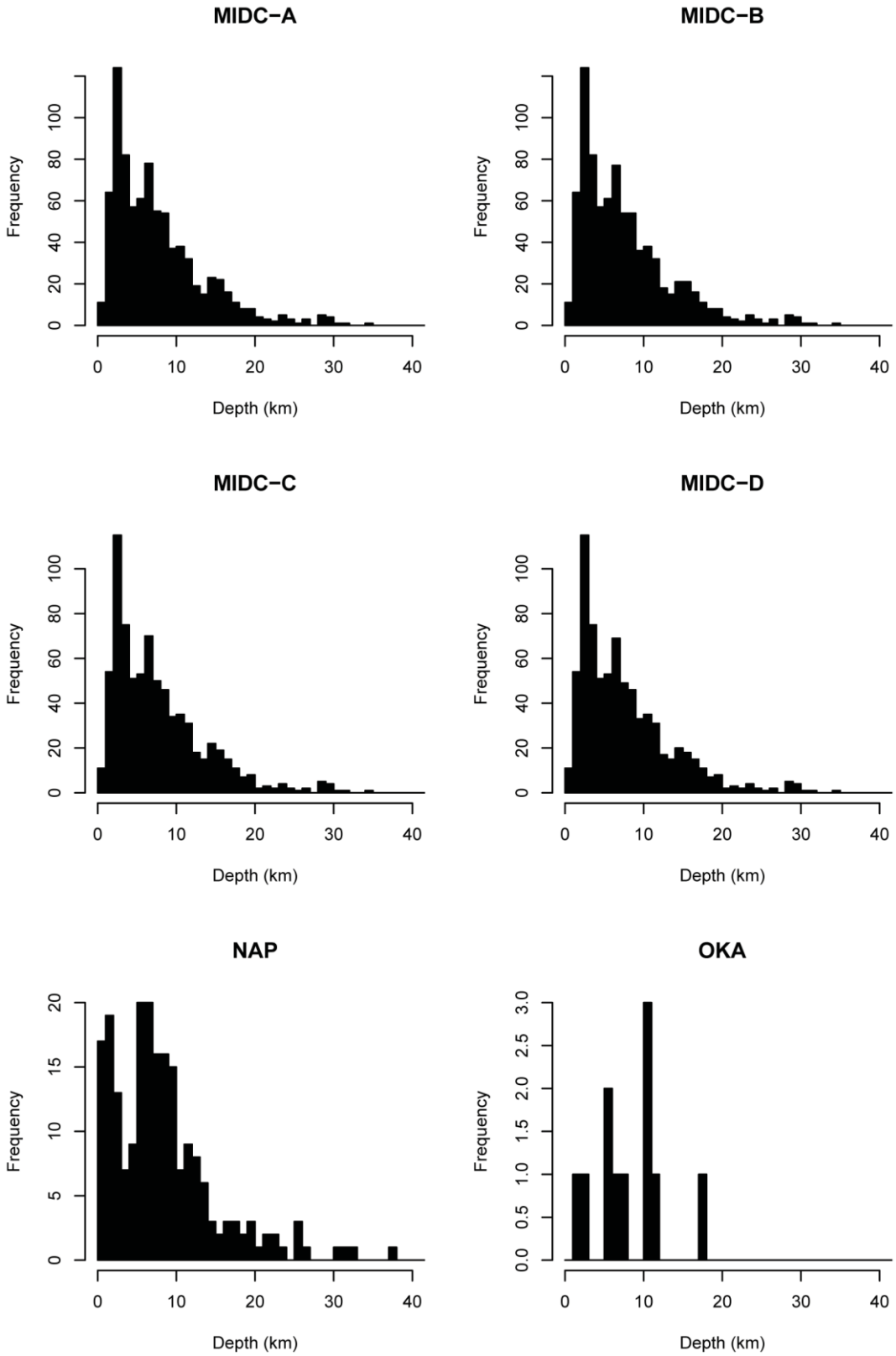

Figure 5.4.4-4 (2 of 3)
Distribution of better-quality focal depths in seismotectonic source zones

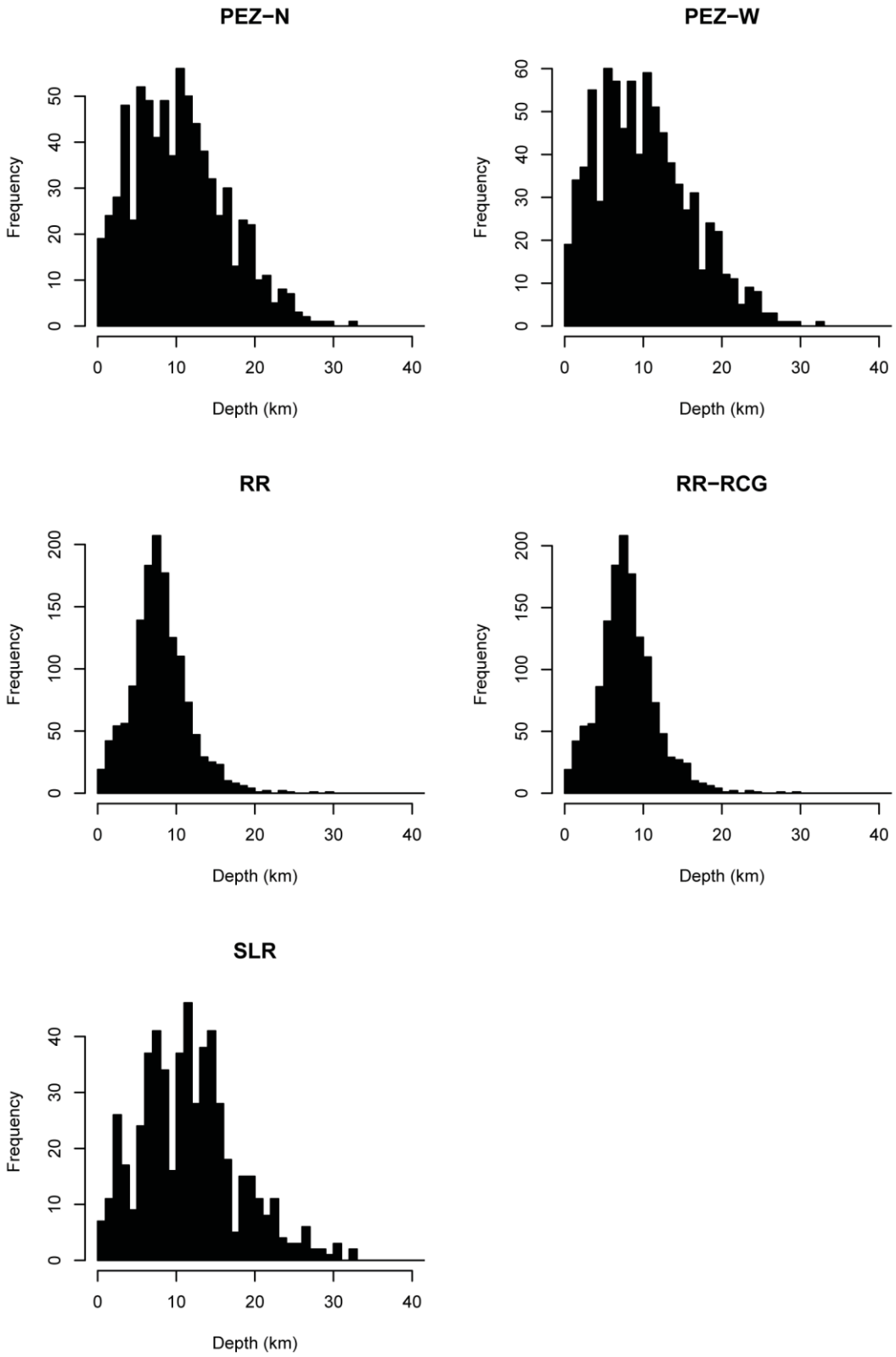

Figure 5.4.4-4 (3 of 3)
Distribution of better-quality focal depths in seismotectonic source zones