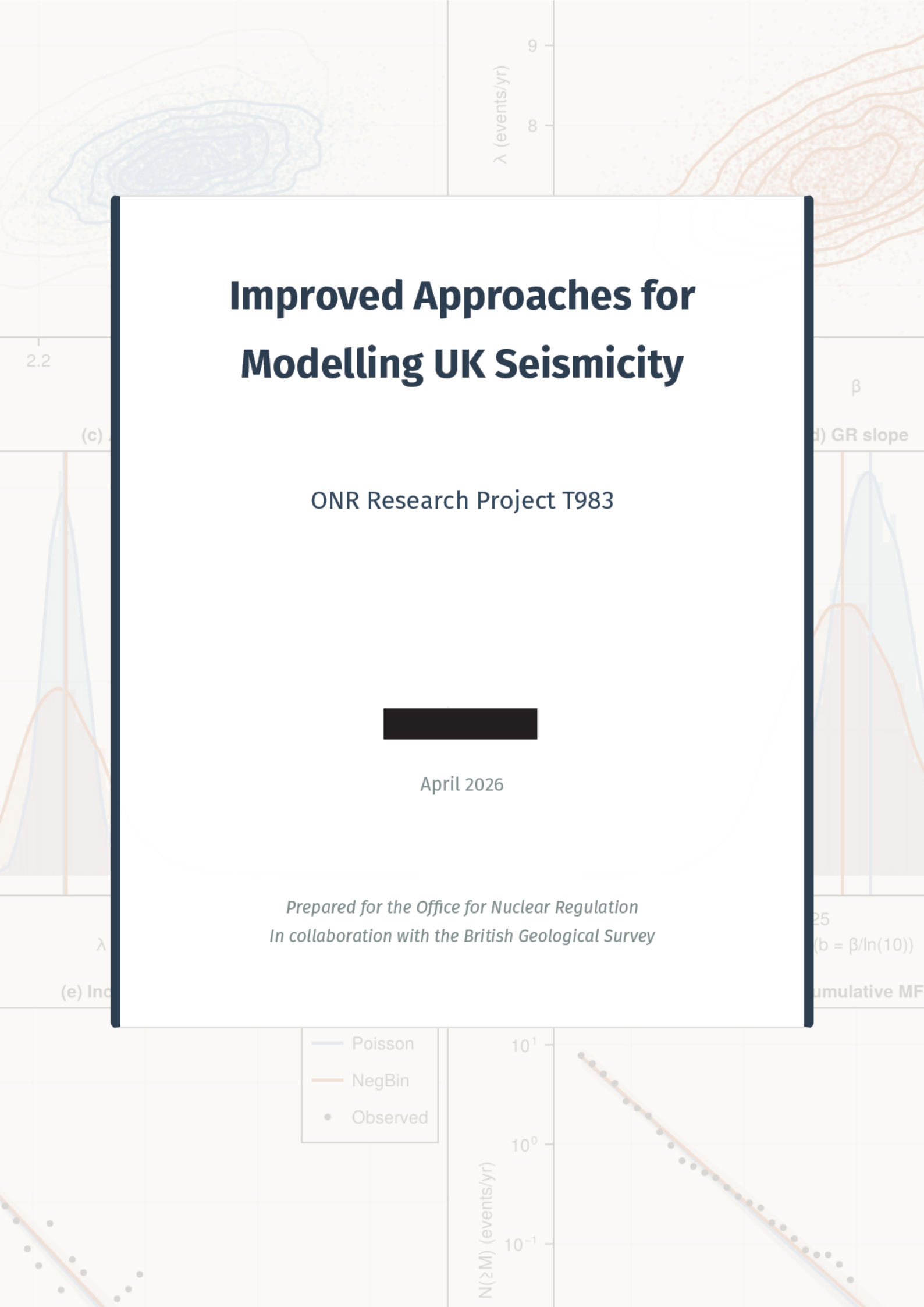

Improved Approaches for Modelling UK Seismicity

ONR Research Project T983

[redacted]

April 2026

Prepared for the Office for Nuclear Regulation
In collaboration with the British Geological Survey

Contents

1 Executive Summary

This report presents improved approaches for calibrating magnitude-frequency distributions used in probabilistic seismic hazard analysis (PSHA) for UK nuclear sites. The work addresses three interrelated themes:

Theme A (50% effort): Optimal calibration of magnitude-frequency parameters, including Bayesian alternatives to penalized maximum likelihood estimation, explicit treatment of magnitude uncertainties and rounding, and investigation of prior influence on parameter estimates.

Theme B (20% effort): Treatment of epistemic uncertainty across area sources, including correlation effects, discrete approximations for logic trees, and conservation of uncertainty under zone subdivision.

Theme C (30% effort): Joint inversion frameworks for multiple area sources using hierarchical models that enable partial pooling of information across zones.

1.1 Principal Findings

Unifying Framework Developed: A "Full Bayesian" Bayesian model has been developed and validated that encompasses all classical methods as special limiting cases:

- **Weichert (1980)**: Recovered when magnitude uncertainty $\sigma \to 0$
- **Johnston PMLM (1994)**: Recovered by adding an informative prior on β
- **Tinti & Mulargia (1985)**: Rate bias formula emerges from the expected count integral
- **NUREG N*/M***: Our marginalisation is the rigorous Bayesian treatment of the same insight

By demonstrating these limiting cases, the Full Bayesian model captures everything previous methods attempted, but integrates all effects consistently within a single coherent framework.

Key technical findings include:

1. **Magnitude uncertainty effects**: The PMLM approach currently used in UK practice ignores magnitude uncertainties entirely, leading to systematic biases that do not average out with more data. The magnitude of these biases depends critically on the buffer between the completeness threshold m_c and the analysis threshold $m_{\min}$: the Tinti-Mulargia formula predicts ~27% activity rate overestimation for typical UK conditions ($\sigma \approx 0.30$, combining local magnitude measurement uncertainty and M_L-to-M_w conversion uncertainty) when no buffer is present, with b-value biases of similar or greater magnitude.

Current UK practice benefits from conservative completeness thresholds that provide a buffer of $\approx 3\sigma$, reducing rate biases to ~10–15% and b-value biases to ~5–10%. However, if the analysis threshold is pushed closer to the completeness limit—as might be tempting to increase sample sizes—these biases grow substantially. The Full Bayesian Bayesian model developed here accounts for magnitude uncertainty explicitly and achieves $< 2\%$ bias regardless of the buffer.

2. **Poisson vs Multinomial**: Multinomial methods (Aki, 1965; Weichert, 1980) treat the event count as fixed, making them structurally unable to handle magnitude uncertainty. Only Poisson-based methods can model the scatter-IN/OUT processes that arise from measurement error. Although PMLM uses a Poisson framework, it does not exploit this capability—it ignores magnitude uncertainty in practice, producing the biases noted above. This distinction, identified by Dyck (1985), is fundamental to proper uncertainty quantification.
3. **Uncertainty calibration**: Monte Carlo validation using 100 synthetic catalogues with known parameter values confirms that the Full Bayesian Bayesian model produces correctly-sized posterior uncertainties. The average posterior standard deviation closely matches the observed scatter across replications (ratio ≈ 0.7–1.1). The nominal 95% credible intervals contain the true parameter values 91% of the time for b and 97% for λ, indicating that the model neither overstates nor materially understates uncertainty.
4. **PMLM prior assessment**: The Johnston PMLM prior mean ($b = 1.0$) is consistent with UK regional estimates in M_w scale and supported by global SCR analogues. However, the associated standard deviation ($\sigma_b \approx 0.087$) acts as a strong prior that substantially suppresses the variability observed across UK source zones. Hierarchical analysis reveals zone-level uncertainties of $\sigma_b \approx 0.12$–0.45, meaning PMLM underestimates epistemic uncertainty by 40–80%. This suppression directly affects hazard fractiles used in regulatory assessments.
5. **Overdispersion in event counts**: UK seismicity exhibits significant overdispersion at all magnitude thresholds tested, strongly favouring a negative binomial over the Poisson distribution assumed in standard practice. This finding is confirmed under both the standard PMLM framework ($\varphi \approx 5$–11) and a hybrid Binned L5 model that retains the full Full Bayesian treatment of magnitude uncertainties, conversion, and selection effects ($\varphi \approx 3$–5). The Binned L5 analysis shows that properly accounting for magnitude uncertainties does not explain away the overdispersion – it reveals more of it. Point estimates (b-value, mean rate) are essentially unaffected, but the Poisson assumption underestimates activity rate uncertainty by factors of 2–4×. This propagates directly to hazard confidence intervals and regulatory fractiles.

6. **Buffer zone importance**: Because PMLM and other standard methods ignore magnitude uncertainty, their accuracy depends heavily on the buffer between the completeness threshold m_c and the analysis threshold $m_{\min}$. With no buffer, Eddington bias can be severe; a buffer of ≈ 1.0 magnitude unit ($\approx 3\sigma$ for typical UK measurement uncertainty) reduces residual bias to ~5%. Current UK NSHM practice achieves approximately this buffer through conservative completeness thresholds – but not necessarily by design. Reducing the analysis threshold to capture more events (a natural temptation for data-sparse zones) erodes this protection and increases bias, making it counterproductive unless methods that explicitly account for magnitude uncertainty are used.
7. **Zone distinguishability and cross-zone correlations**: 68% of UK source zones are statistically indistinguishable from the regional b-value, with apparent differences largely attributable to sparse data. Hierarchical modelling reveals that the shared PMLM prior induces strong positive correlations between zone b-values ($\rho(b_i, b_j) \approx 0.94$ under the standard $W = 25$ specification), while physical moment release constraints and boundary uncertainty introduce negative correlations. Standard practice treats zone parameters as independent in logic trees, which is inconsistent with all three of these correlation sources and leads to incorrect propagation of epistemic uncertainty through hazard calculations. Two practical remedies are identified: post-hoc importance sampling, which re-weights independently-derived logic tree branches using moment release constraints to induce physical correlations while preserving existing workflows; and hierarchical Bayesian models, which capture the correlation structure directly and enable partial pooling of information across data-sparse zones.
8. **Logic tree discretization**: Current UK practice constructs 3×3 or 5×5 grids of seismicity parameters and assigns weights proportional to the renormalised likelihood function. This approach has no mathematical basis: it does not preserve the moments of the joint parameter distribution, nor the fractiles of predicted exceedance rates or hazard estimates. There is no justification for its continued use. Mathematically grounded alternatives (e.g. Miller-Rice, eigenvalue-based schemes) do preserve moments, but still produce 7–8% errors in the 84th percentile of exceedance rates. An optimised Heavy-Tail scheme reduces this to 1.5%, with further improvement to 0.5% using asymmetric grids.
9. **Moment release constraints**: The UK NSHM 2020 implies moment release broadly consistent with geodetic constraints (1.57×10^{16} vs 1.42×10^{16} N·m/year from Main et al. (1999)). However, when zone parameters are sampled independently in a logic tree, the implied range of total moment release spans orders of magnitude—far exceeding what geophysical observations permit. Moment release constraints provide a physically-motivated basis for inducing correlations between zone parameters and pruning implausible logic tree branches, as discussed above under zone distinguishability and cross-zone

correlations. Parameter elasticities show $M_{\max}$ has 8–10$\times$ greater impact on moment release than λ or b, highlighting the importance of $M_{\max}$ characterisation in any moment-constrained framework.

1.2 Overall Assessment

Taken together, these findings demonstrate that current UK practice for seismicity parameter estimation contains multiple systematic deficiencies that compound through the hazard calculation. Biased point estimates from ignoring magnitude uncertainty, suppressed epistemic uncertainty from overly informative priors, underestimated rate variability from the Poisson assumption, ignored cross-zone correlations, and mathematically unjustified logic tree discretization each contribute to hazard estimates that are biased in both the mean and the fractiles used for regulatory decision-making. The magnitude of these effects is non-trivial: individually, the identified biases range from 5–30% on parameter estimates and 40–80% on uncertainty measures. This report provides concrete, implementable alternatives for each of these issues, offering a clear path toward more defensible seismic hazard assessments.

2 Introduction

2.1 Background

Seismic hazard often contributes to the bounding load case for the design of Structures, Systems and Components (SSCs) at nuclear sites. The external hazards Safety Assessment Principle (SAP) EHA.4 expects design basis events for natural hazards to be characterized by frequency of exceedance hazard curves defined as a 1 in 10,000 years event.

Probabilistic seismic hazard analysis (PSHA) requires two key inputs: the Ground Motion Model (GMM) and the Seismic Source Model (SSM). This research focuses on the SSM, specifically the statistical analysis to determine seismicity parameters for area sources.

The primary output of a PSHA is a hazard curve defining the average annual rate of exceeding different levels of an intensity measure (like spectral acceleration at a given period and damping ratio). While there a number of ways to represent this calculation (Baker et al., 2021), a simple expression is:

$$\lambda(IM \geq im) = \sum_{i=1}^{n_{\text{sources}}} \sum_{j=1}^{n_{\text{rup},i}} P\big(IM \geq im \mid \text{rup}_j, \text{site}\big)\lambda\big(\text{rup}_j\big) \tag{2.1}$$

where: n_{sources} is the number of seismic sources, $n_{\text{rup},i}$ is the number of rupture scenarios for source i, $P\big(IM \geq im \mid \text{rup}_j, \text{site}\big)$ is the probability of exceeding intensity measure IM given rupture scenario j and site, and $\lambda\big(\text{rup}_j\big)$ is the annual rate of occurrence of rupture scenario j. In the UK, it is common to use area sources, and to represent the magnitude-frequency distribution of earthquakes in each source zone using the doubly-bounded exponential (Gutenberg-Richter) distribution (Baker et al., 2021; Cornell & Vanmarcke, 1969; Gutenberg & Richter, 1944). To first-order (ignoring issues related to kinematic admissability, and boundary constraints, *etc*), the rate of occurrence of a ruptture scenario is equivalent to the rate of occurrence of earthquakes of the corresponding magnitude in the source zone.

If an area source was subdivided into multiple smaller zones, the sum of the rates from the sub-zones should equal the rate from the original zone. So, if we define the magnitude-frequency distribution in each zone in terms of an activity rate density (events/year/km²) rather than an absolute activity rate (events/year), then the rates will sum correctly when zones are subdivided. Under this framework, a rupture scenario is characterised not only by its magnitude m_j but also by its location within the source zone and the geometry of the rupture (strike, dip, focal depth). The total rate of earthquakes at magnitude m_j in a source zone is partitioned across all possible rupture locations and orientations, so that

each scenario rup_j has a rate $\lambda(\mathrm{rup}_j)$ that accounts for its share of the total. Integrating over all rupture positions and geometries recovers the total activity rate at that magnitude:

$$\lambda(M = m_j) = \int_A \int_\Omega \lambda(\mathrm{rup}_j \mid \boldsymbol{x}, \omega) d\omega dA \tag{2.2}$$

where $\boldsymbol{x}$ denotes position within the source zone, ω represents the rupture geometry (strike, dip, focal depth, relative focal position within the rupture, *etc*), and $\lambda(M = m_j)$ is the activity rate density at magnitude m_j (events/year/km²). For area sources with uniform seismicity, the spatial and geometric components are often simplified by assuming uniform distributions over the zone area and the admissible focal mechanism space.

The activity rate density is typically modelled using the Gutenberg-Richter distribution (see Section 3 for details):

$$\lambda(M = m) = \lambda_n \cdot f_M(m \mid \beta, m_n, m_x) \tag{2.3}$$

where $f_M(m \mid \beta, m_n, m_x)$ is the probability density function defined in Equation 3.1, and λ_n is the activity rate (events/year/km²) – the number of events per year, per square kilometre, with a magnitude of at least the minimum magnitude m_n. The domain of support for the distribution is bounded between a minimum magnitude m_n and a maximum magnitude m_x, and $\beta = b \cdot \ln(10)$ where b is the more commonly referred to *b*-value.

The maximum magnitude m_x cannot typically be estimated through statistical analyses of the earthquake catalogue in places like the UK, so it is usually set based on geological and tectonic considerations, and regional tectonic analogues. To specify the magnitude-frequency distribution for an area source, we therefore need to estimate two parameters from the earthquake catalogue: the activity rate λ_n and the *b*-value (or equivalently β). The parameter m_n must also be specified for use in PSHA calculations, but this is a modelling decision that does not relate to the statistical estimation problem. However, when analysing earthquake catalogues, we must also consider the minimum magnitude $m_{\min}$ – which is related to the lowest magnitude above which we believe the catalogue is complete for a given time period. In the UK, $m_{\min}$ is often set to around 3.0, while m_n in the PSHA is around 4.0-4.5 (Mosca et al., 2022).

2.2 Research Themes

This research addresses two broad, interrelated topics for seismicity analysis in the UK: (1) the optimal calibration of parameters for magnitude-frequency distributions; and (2) the treatment of epistemic uncertainty over a collection of area sources. As these are interrelated, their combined consideration leads to a third area of investigation – how to perform joint inversions for seismic source models.

These issues are not exclusive to the UK; they arise in most geographical regions. However, their importance is exacerbated in low-seismicity regions like the UK where parametric uncertainty in the estimation of seismicity parameters is large, and where historical approaches to seismic source zonation have favoured relatively large numbers of discrete area sources.

The research is organised around three themes, with indicative effort allocations:

Theme A — Optimal Calibration of Magnitude-Frequency Parameters (50% effort): This theme investigates alternatives to the penalised maximum likelihood method, including complete Bayesian frameworks with formal prior specification. Key sub-topics include:

- Explicit treatment of magnitude uncertainties and rounding
- Investigation of posterior distributions and their effects on rate estimates
- Development of a "Full Bayesian" Bayesian model that addresses the limitations of current methods
- Validation through synthetic data simulation

Theme B — Treatment of Epistemic Uncertainty (20% effort): UK hazard studies typically define a relatively large number of area source zones, leading to low event counts per zone, increased parametric uncertainty, and large logic trees that may include physically implausible parameter combinations. This theme addresses:

- Conservation of epistemic uncertainty when sources are subdivided
- Correlation of logic tree branches for consistent parameter sets across the seismic source model
- Optimal discretization schemes for logic tree construction
- External constraints such as seismic moment release to ensure physical plausibility

Theme C — Joint Inversion for Multiple Area Sources (30% effort): Combining the previous themes leads to consideration of joint inversion of seismicity parameters for all source zones simultaneously, rather than obtaining parameters conditioned on assumed source boundaries. This theme explores:

- Hierarchical (multi-level) models enabling partial pooling across zones
- Correlation structure of seismicity parameters across area sources
- Integration of external constraints (moment release, geodetic data) into the inversion framework
- Exploratory analysis of data-driven source zonation as a counterfactual to current expert-defined boundaries

2.3 Current UK Practice

UK hazard studies conventionally calibrate the parameters of the doubly-bounded exponential (Gutenberg-Richter) magnitude-frequency distribution using the penalized maximum likelihood estimation method described in Johnston et al. (1994) – often referred to as the PMLM in UK practice. While this approach has strong precedent, it has known limitations:

- Does not account for individual magnitude uncertainties
- Does not handle 'soft' incompleteness levels
- Ignores rounding of magnitude estimates
- Uses ad-hoc weighting rather than formal prior specification

Because this approach is the *de facto* standard for UK seismicity analysis, it is important to understand its limitations and to develop improved methods that address them. Before making those recommendations, it is instructive to look closely at the behaviour of the PMLM, and Section 4 provides extensive details on this method that are not readily found elsewhere; the original report is not easily accessible (Johnston et al., 1994).

2.4 Research Objectives

This project develops improved methods addressing the limitations of current practice:

1. Establish a rigorous Bayesian framework that properly handles magnitude uncertainty, selection effects, and variable completeness
2. Demonstrate how classical methods emerge as limiting cases of this framework
3. Validate methods through synthetic data testing
4. Apply to UK earthquake catalogues and source zones
5. Develop recommendations for improved practice

2.5 Scope and Organization

The remainder of this report is organised as follows. Section 3 provides the basic mathematical definitions for the Gutenberg-Richter distribution. Section 4 then provides an extensive treatment of the penalised maximum likelihood method (PMLM) that is currently used in UK practice—this detailed coverage is provided because the original source (Johnston et al., 1994) is not readily accessible. Section 5 describes the UK earthquake catalogue and source zone definitions used throughout this work.

The main technical contributions are then organised by theme:

Theme A is covered in three sections: Section 6 presents the Full Bayesian Bayesian model; Section 7 addresses magnitude uncertainties and validation; and Section 8 examines statistical issues in magnitude conversions.

Theme B is covered in three sections: Section 9 addresses correlation between parameters and logic tree discretization; Section 10 presents regional analysis and zone correlations; and Section 11 develops seismic moment release constraints for ensuring physical plausibility.

Theme C is covered in two sections: Section 12 presents the hierarchical model framework for joint inversion; and Section 13 explores data-driven approaches to source zonation, including K-means clustering and merging of statistically indistinguishable zones.

Section 14 addresses alternative occurrence models, specifically testing the Poisson assumption against the negative binomial distribution. This analysis is relevant to the project scope item on alternative magnitude-frequency distributions – while the GR distribution is retained, the Poisson assumption for event counts is relaxed to allow for overdispersion. We note that many other magnitude-frequency distributions have been proposed as alternatives to the GR distribution. These are not limited to historical reviews from decades ago (Main, 1996; Utsu, 1999), but also represent approaches adopted for recent high-profile European seismic hazard and risk projects (Bourne & Oates, 2020; Danciu et al., 2024). However, a full analysis of these alternative magnitude-frequency distributions is beyond the scope of the current project as the focus is upon enhancing current UK practice (that overwhelmingly uses the Gutenberg-Richter distribution).

Finally, Section 15 draws together the findings across all three themes and provides recommendations for practice.

3 Gutenberg-Richter Distribution

The Gutenberg-Richter (GR) distribution, as originally defined (Gutenberg & Richter, 1944), is never actually used in practice. Rather, the doubly-bounded exponential distribution is used, which accounts for the fact that the domain of support for the distribution is limited at lower, m_n, and upper magnitudes, m_x, by practical and physical constraints (Cornell & Vanmarcke, 1969). The relationship between these distributions is explicitly defined in Baker et al. (2021).

3.1 Distribution Functions

The probability density function (pdf) of the doubly-bounded exponential distribution (hereafter referred to as the GR distribution) is:

$$f_M(m \mid \beta, m_n, m_x) = \frac{\beta e^{-\beta(m-m_n)}}{1 - e^{-\beta(m_x-m_n)}} \tag{3.1}$$

where $\beta = b \cdot \ln(10)$, m_n is the minimum magnitude, and m_x is the maximum magnitude (Baker et al., 2021). The parameter b is the Gutenberg-Richter b-value, a measure of the slope of the magnitude-frequency distribution, defined in terms of $\log_{10}$ numbers of events for a given magnitude (Gutenberg & Richter, 1944).

The cumulative distribution function is:

$$F_M(m \mid \beta, m_n, m_x) = \frac{1 - e^{-\beta(m-m_n)}}{1 - e^{-\beta(m_x-m_n)}} \tag{3.2}$$

The mean magnitude is:

$$\overline{m} = m_n + \frac{1}{\beta} - \frac{m_x - m_n}{e^{\beta(m_x-m_n)} - 1} \tag{3.3}$$

and the original MLE solution of Aki (1965) corresponds to the case where $m_x \gg m_n$ so that:

$$\hat{\beta}_{\text{Aki}} = \frac{1}{\overline{m} - m_n} \tag{3.4}$$

While this estimator is no longer used (as we will see throughout the report), its still instructive to appreciate that when we have an earthquake catalogue that is complete for events above m_n, the slope (β and hence b) is tied to how far above m_n the magnitude in the sample is. The smaller $\overline{m} - m_n$, the higher β, and *vice versa*.

3.2 Activity Rates and Exceedance

The distribution functions defined above describe only the *shape* of the magnitude distribution—i.e., the relative likelihood of different magnitudes given that an event occurs. For seismic hazard analysis, we must also specify the *rate* at which events occur. This requires introducing the activity rate parameter λ_n, which represents the mean annual rate of earthquakes with magnitude $M \geq m_n$ in the source zone.

The complete parameterisation of the magnitude-frequency relation therefore requires three quantities:

- λ_n: the activity rate (events per year above m_n)
- β: the slope parameter controlling the relative frequency of small versus large events
- m_x: the maximum magnitude (often treated as known or estimated separately)

Together with the minimum magnitude m_n (which is typically a modelling choice rather than an estimated parameter), these define the complete magnitude-frequency model.

3.2.1 Rate of Exceedance

The rate at which earthquakes exceed a given magnitude threshold m is obtained by combining the activity rate with the complementary CDF (survival function):

$$\lambda(M \geq m) = \lambda_n \cdot [1 - F_M(m)] = \lambda_n \cdot \frac{e^{-\beta(m-m_n)} - e^{-\beta(m_x-m_n)}}{1 - e^{-\beta(m_x-m_n)}} \tag{3.5}$$

At the minimum magnitude, this reduces to $\lambda(M \geq m_n) = \lambda_n$, confirming the interpretation of λ_n as the activity rate above m_n. At the maximum magnitude, $\lambda(M \geq m_x) = 0$.

For many practical purposes, and particularly when $m_x - m_n$ is large (typically 3–4 magnitude units), the exponential terms involving m_x become negligible and Equation 3.5 simplifies to:

$$\lambda(M \geq m) \approx \lambda_n \cdot e^{-\beta(m-m_n)} \tag{3.6}$$

This is the classical Gutenberg-Richter relation, often written in terms of the b-value as:

$$\log_{10} \lambda(M \geq m) = a - bm \tag{3.7}$$

where $a = \log_{10} \lambda_n + bm_n$ is the intercept parameter. While the (a, b) parameterisation is common in seismological literature, the (λ_n, β) parameterisation used here is preferable for statistical estimation because λ_n has a direct physical interpretation as an observable rate.

3.2.2 Rate of Occurrence

For probabilistic seismic hazard analysis, the fundamental quantity is the rate of occurrence of earthquakes at a specific magnitude (or within a narrow magnitude range). This is obtained by combining the activity rate with the PDF:

$$\lambda(M = m) = \lambda_n \cdot f_{M(m)} = \lambda_n \cdot \frac{\beta e^{-\beta(m-m_n)}}{1 - e^{-\beta(m_x-m_n)}} \tag{3.8}$$

This represents the activity rate *density*—the rate of events per unit magnitude. For a small magnitude interval $[m, m + \delta m)$, the expected number of events per year is approximately $\lambda(M = m) \cdot \delta m$.

The occurrence rate density is the quantity that ultimately enters the hazard integral in Equation 2.1. As discussed in Section 2, the total rate at a given magnitude in a source zone is partitioned across all possible rupture scenarios—each defined by a location and rupture geometry. For area sources with uniform seismicity and uniform focal mechanism distributions, the rate of each rupture scenario at magnitude m_j is simply:

$$\lambda(\mathrm{rup}_j) = \frac{\lambda(M = m_j)}{A_{\mathrm{source}} \cdot N_\Omega} \tag{3.9}$$

where A_{source} is the zone area and N_Ω is the number of discrete rupture geometries considered. Summing over all locations and orientations recovers $\lambda(M = m_j)$, ensuring consistency with Equation 3.8.

3.2.3 Relationship Between Parameters

The activity rate λ_n and slope β play distinct but complementary roles:

- λ_n scales the overall rate of seismicity (and hence hazard) without changing the shape of the magnitude distribution
- β controls how rapidly the occurrence rate decreases with magnitude (and influences which scenarios contribute to the hazard for a given return period)

An important consequence is that these parameters are not independent in their effect on hazard. At any magnitude $m > m_n$, the occurrence rate depends on both parameters through Equation 3.8. This coupling, combined with the constraint that the observed total number of events must be consistent with the model, induces correlations that must be preserved when propagating uncertainties through hazard calculations.

3.2.4 Rate Transformation and Correlation Structure

A subtle but practically important issue arises when transforming between rates at different reference magnitudes. Methods such as PMLM (see Section 4) fit the parameters λ_n and β, where λ_n is the rate at the catalogue minimum magnitude m_n. However, hazard calculations typically use a higher minimum magnitude $m_{\min} > m_n$. The rate at $m_{\min}$ is:

$$\lambda(m_{\min}) = \lambda_n \cdot e^{-\beta(m_{\min} - m_n)} = \lambda_n \cdot e^{-\beta \Delta m} \tag{3.10}$$

where $\Delta m = m_{\min} - m_n > 0$ is the magnitude shift.

An important question to consider is: how does the correlation with β change under this transformation? The reason that this is important is that logic tree branches are normally specified to reflect the correlation between the GR parameters λ_n and β, but the hazard

calculations use $\lambda(m_{\min})$. The issue is therefore, how does the correlation between $\lambda(m_{\min})$ and β relate to the correlation between λ_n and β?

3.2.4.1 Derivation of Transformed Correlation

Working in log-space simplifies the analysis. Define:

$$L = \ln(\lambda_n), \quad W = \ln[\lambda(m_{\min})] = L - \beta\Delta m \tag{3.11}$$

The covariance between W and β is:

$$\begin{aligned} \mathrm{Cov}(W, \beta) &= \mathrm{Cov}(L - \beta\Delta m, \beta) \\ &= \mathrm{Cov}(L, \beta) - \Delta m \cdot \mathrm{Var}(\beta) \end{aligned} \tag{3.12}$$

This can be written in terms of the original correlation $\rho_{L\beta} = \mathrm{Cov}(L, \beta)/\left(\sigma_L \sigma_\beta\right)$:

$$\mathrm{Cov}(W, \beta) = \rho_{L\beta}\sigma_L\sigma_\beta - \Delta m\sigma_\beta^2 \tag{3.13}$$

The correlation between $W = \ln[\lambda(m_{\min})]$ and β is therefore:

$$\rho_{W\beta} = \frac{\mathrm{Cov}(W, \beta)}{\sigma_W \sigma_\beta} \tag{3.14}$$

To find σ_W, we use the variance transformation:

$$\begin{aligned} \mathrm{Var}(W) &= \mathrm{Var}(L - \beta\Delta m) \\ &= \mathrm{Var}(L) + (\Delta m)^2\ \mathrm{Var}(\beta) - 2\Delta m\ \mathrm{Cov}(L, \beta) \\ &= \sigma_L^2 + (\Delta m)^2\sigma_\beta^2 - 2\Delta m\rho_{L\beta}\sigma_L\sigma_\beta \end{aligned} \tag{3.15}$$

Substituting Equation 3.13 and Equation 3.15 into Equation 3.14:

$$\rho_{W\beta} = \frac{\rho_{L\beta}\sigma_L - \Delta m\sigma_\beta}{\sqrt{\sigma_L^2 + (\Delta m)^2\sigma_\beta^2 - 2\Delta m\rho_{L\beta}\sigma_L\sigma_\beta}} \tag{3.16}$$

3.2.4.2 Sign Change Condition

The sign of $\rho_{W\beta}$ is determined by the numerator. The correlation changes from positive to negative when:

$$\rho_{L\beta}\sigma_L - \Delta m\sigma_\beta < 0 \tag{3.17}$$

Rearranging, the critical magnitude shift, Δm_{crit}, at which the correlation changes sign is:

$$\Delta m_{\text{crit}} = \rho_{L\beta} \cdot \frac{\sigma_L}{\sigma_\beta} \tag{3.18}$$

For $\Delta m > \Delta m_{\text{crit}}$, the correlation $\rho(\ln\lambda(m_{\min}), \beta)$ becomes negative even if the original correlation $\rho(\ln\lambda_n, \beta)$ was positive.

3.2.4.3 Practical Implications for UK Practice

Consider typical UK values from the PMLM analysis:

- $\rho_{L\beta} \approx 0.4$ to 0.5 (positive correlation at m_n)
- $\sigma_L \approx 0.15$ to 0.25 (corresponding to 15-25% coefficient of variation in λ_n)

- $\sigma_\beta \approx 0.10$ to 0.15 (corresponding to $\sigma_b \approx 0.04$ to 0.065)

The critical shift is approximately:

$$\Delta m_{\text{crit}} \approx \frac{0.45 \times 0.20}{0.12} \approx 0.75 \text{ magnitude units} \tag{3.19}$$

This has significant practical implications:

Rate-Correlation Trade-off:

If the catalogue is fitted down to $m_n = 2.5$ but hazard calculations use $m_{\min} = 4.0$ (a common scenario), then $\Delta m = 1.5 > \Delta m_{\text{crit}}$, and the correlation between $\ln \lambda(m_{\min})$ and β will be **negative** even though the fitted correlation between $\ln \lambda_n$ and β was **positive**.

For logic tree construction: When specifying correlated branches for (λ, β) at the hazard reference magnitude, practitioners must use the **transformed** correlation from Equation 3.16, not the correlation returned directly by the PMLM fit. Or, they must ensure that their method to extrapolate from m_n to $m_{\min}$ results in the appropriate correlation structure being preserved.

3.2.4.4 Physical Interpretation

The sign change occurs because two competing effects are at play:

1. **Positive contribution from λ_n**: As β increases, the PMLM solution requires λ_n to increase (to maintain the observed count N given the shift in $\overline{m}$ toward lower magnitudes with shorter observation periods). This contributes positively to $\text{Cov}(\lambda(m_{\min}), \beta)$.
2. **Negative contribution from the exponential decay**: The factor $e^{-\beta \Delta m}$ decreases as β increases. For large Δm, this effect dominates, pulling the overall correlation negative.

The net correlation reflects the balance between these effects. At small Δm, the positive contribution from λ_n dominates. At large Δm, the negative contribution from the exponential decay dominates.

3.2.4.5 Numerical Example

To illustrate, consider a zone with fitted parameters:

- $\lambda_n = 2.5$ events/year at $m_n = 2.5$
- $\beta = 2.3$ (i.e., $b = 1.0$)
- $\sigma_{\ln \lambda_n} = 0.20$, $\sigma_\beta = 0.12$
- $\rho(\ln \lambda_n, \beta) = +0.45$

For hazard calculations at $m_{\min} = 4.5$ (so $\Delta m = 2.0$):

From Equation 3.16:

$$\rho(\ln \lambda(4.5), \beta) = \frac{0.45 \times 0.20 - 2.0 \times 0.12}{\sqrt{0.20^2 + 2.0^2 \times 0.12^2 - 2 \times 2.0 \times 0.45 \times 0.20 \times 0.12}} \tag{3.20}$$

$$= \frac{0.090 - 0.240}{\sqrt{0.040 + 0.058 - 0.043}} = \frac{-0.150}{\sqrt{0.055}} = -0.64 \tag{3.21}$$

The correlation has indeed flipped from $+0.45$ to -0.64.

The rate at $m_{\min} = 4.5$ is:

$$\lambda(4.5) = 2.5 \times e^{-2.3 \times 2.0} = 2.5 \times 0.010 = 0.025 \text{ events/year} \tag{3.22}$$

The transformed variance of $\ln \lambda(4.5)$ follows from Equation 3.15, and is the same quantity that appears under the square root in the denominator above:

$$\sigma_{\ln \lambda(4.5)} = \sqrt{\text{Var}(\ln \lambda(4.5))} = \sqrt{\sigma^2_{\ln \lambda_n} + (\Delta m)^2 \sigma^2_\beta - 2\Delta m \rho \sigma_{\ln \lambda_n} \sigma_\beta} = \sqrt{0.055} = 0.23 \tag{3.23}$$

3.2.4.6 Recommendation for Logic Tree Construction

When constructing logic trees for hazard calculations:

1. Fit the GR parameters at the catalogue minimum magnitude m_n to obtain $\left(\hat{\lambda}_n, \hat{\beta}, \boldsymbol{\Sigma}\right)$
2. Transform to the hazard reference magnitude $m_{\min}$ using Equation 3.10
3. Compute the transformed covariance structure using Equation 3.12 and Equation 3.15
4. Use the transformed correlation (which may be negative) when specifying correlated branches

Failure to account for this transformation will result in incorrectly specified logic tree correlations, potentially leading to biased hazard uncertainty estimates.

3.2.4.7 Demonstration of the Impact

To illustrate the practical consequences of ignoring the correlation transformation, we present a series of analyses using illustrative parameter values typical of UK seismicity studies. Consider a source zone with:

- Catalogue minimum magnitude $m_n = 2.5$
- Hazard minimum magnitude $m_{\min} = 4.0$ (i.e., $\Delta m = 1.5$)
- Fitted activity rate $\lambda_n = 3.0$ events/year
- $\beta = 2.3$ ($b = 1.0$)
- Uncertainties: $\sigma_{\ln \lambda_n} = 0.22$, $\sigma_\beta = 0.14$
- Correlation at m_n: $\rho(\ln \lambda_n, \beta) = +0.42$

Figure 3.1 shows how the joint distribution of $(\ln \lambda, \beta)$ changes when transforming from m_n to $m_{\min}$. The left panel shows the distribution at the catalogue minimum, where the positive correlation is evident in the tilted ellipse. The right panel shows the transformed distribution at the hazard reference magnitude – the correlation has flipped to negative, and the uncertainty in $\ln \lambda$ has increased.

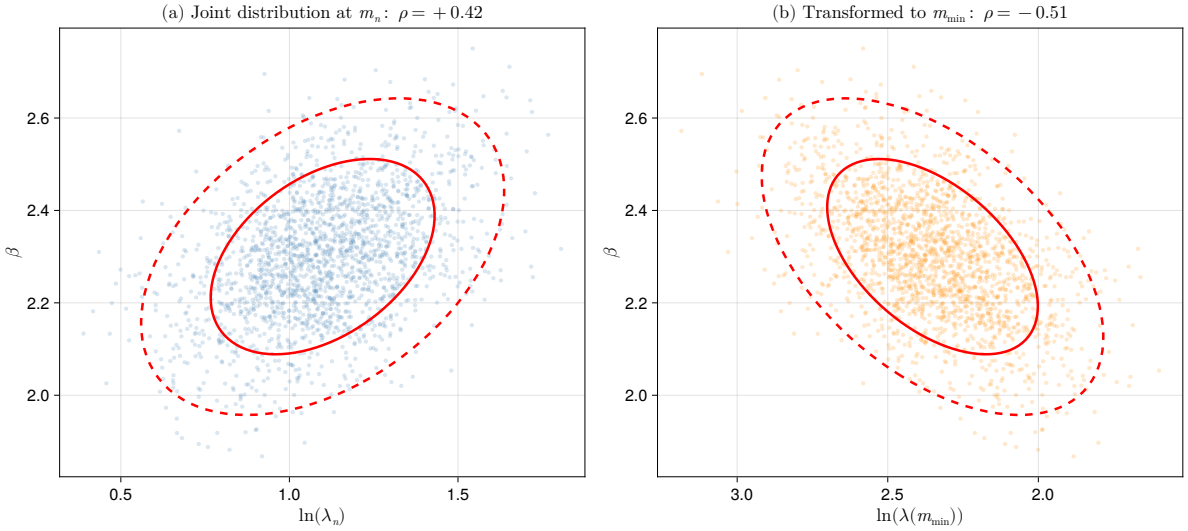

Figure 3.1: Joint distributions of $(\ln \lambda, \beta)$ at different reference magnitudes. Left: at catalogue minimum $m_n = 2.5$, showing positive correlation ($\rho = +0.42$). Right: at hazard minimum $m_{\min} = 4.0$, showing negative correlation after transformation. The correlation sign change occurs because the magnitude shift $\Delta m = 1.5$ exceeds the critical value $\Delta m_{\text{crit}} \approx 0.7$.

Figure 3.2 shows how the correlation varies continuously as a function of the magnitude shift. For this parameter set, the critical shift $\Delta m_{\text{crit}} \approx 0.7$ magnitude units marks the crossover from positive to negative correlation. The common practice of using $m_{\min} = 4.0$ or higher for hazard calculations places practitioners firmly in the negative correlation regime.

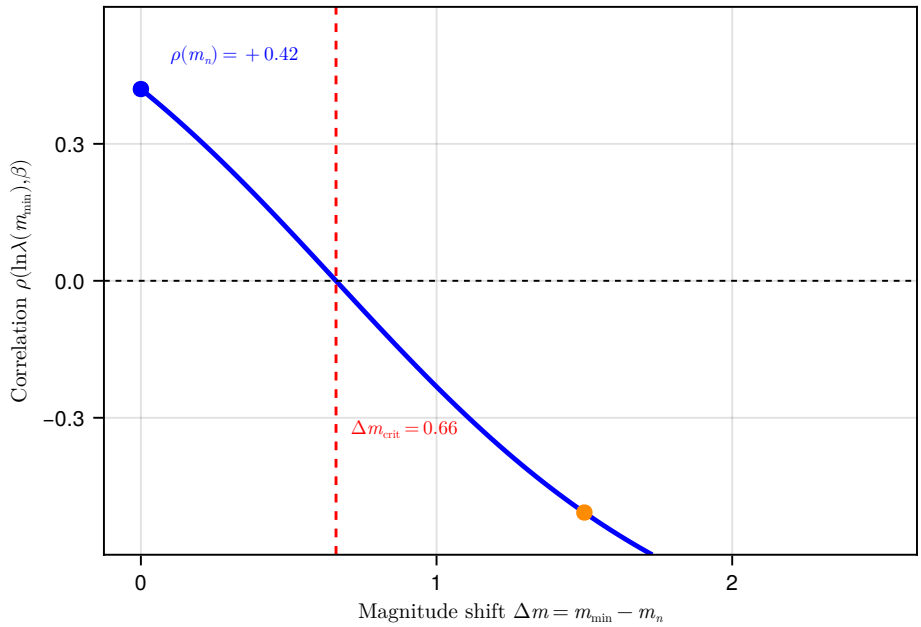

Figure 3.2: Correlation between $\ln \lambda$ and β as a function of the magnitude shift Δm. The correlation transitions from positive at small shifts to negative for shifts exceeding $\Delta m_{\text{crit}} \approx 0.7$. Typical UK practice uses $\Delta m = 1.0$ to 2.0, placing the hazard reference correlation firmly in the negative regime.

The most important practical consequence concerns the uncertainty in exceedance rates. Figure 3.3 compares three approaches:

1. **Monte Carlo reference**: Direct sampling from the correctly correlated joint distribution at m_n, then transforming each sample to compute rates at higher magnitudes
2. **Correct correlation**: Using the transformed (negative) correlation at $m_{\min}$
3. **Incorrect correlation**: Naively using the original (positive) correlation at $m_{\min}$

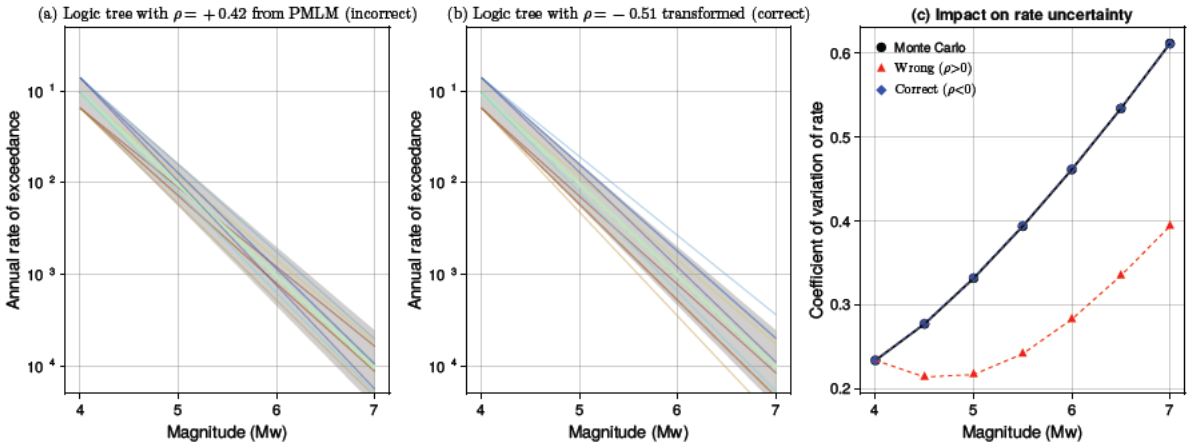

Figure 3.3: Impact of correlation sign on magnitude-frequency uncertainty bands. Left: Using the correctly transformed (negative) correlation produces uncertainty that grows appropriately with magnitude. Right: Using the incorrect (positive) correlation dramatically underestimates uncertainty at higher magnitudes. The coefficient of variation (CoV) comparison shows errors of 35–40% at magnitudes relevant for hazard analysis.

Critical Finding: Uncertainty Underestimation

Using the fitted correlation at m_n directly for hazard calculations at $m_{\min}$ (without transformation) can underestimate the coefficient of variation in exceedance rates by 35–40% at magnitudes in the M_w 5.5–6.5 range that often dominate seismic hazard. This underestimation becomes more severe as the gap between m_n and $m_{\min}$ increases.

Figure 3.4 illustrates how this error manifests in logic tree construction. When building a 3×3 grid of (λ, β) branches, the joint weights should reflect the correlation between parameters. With negative correlation, high-λ/low-β and low-λ/high-β combinations receive enhanced weight, while high-λ/high-β and low-λ/low-β combinations are down-weighted. Using the wrong correlation sign reverses this pattern entirely.

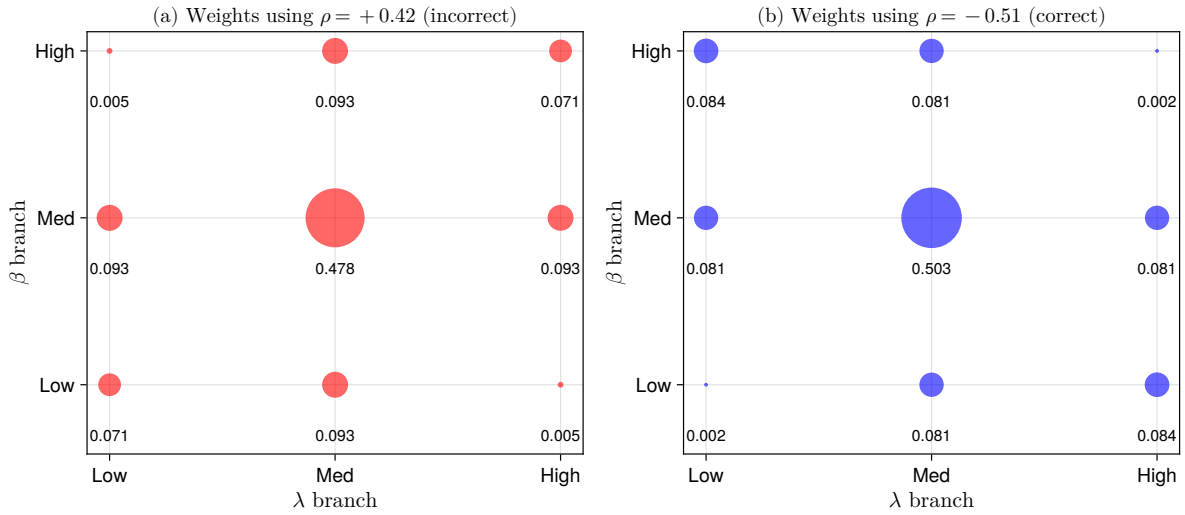

Figure 3.4: Logic tree weights for the same 3×3 grid of (λ, β) nodes evaluated under two different correlations. Left: weights using the incorrect positive correlation ($\rho = +0.42$) from PMLM at m_n. Right: weights using the correct negative correlation ($\rho = -0.51$) transformed to $m_{\min}$. With positive correlation, the high-λ/high-β and low-λ/low-β corners receive enhanced weight. With negative correlation, this reverses: high-λ/low-β and low-λ/high-β combinations dominate, while the opposite corners become negligible (weight ≈ 0.002). Using the wrong correlation sign assigns probability to the wrong regions of parameter space.

This analysis demonstrates that the sign of correlation matters fundamentally for hazard uncertainty quantification. Practitioners using methods like PMLM must transform the fitted covariance structure to the hazard reference magnitude before constructing logic trees. The transformation is straightforward (Equation 3.12 and Equation 3.15), but failure to apply it will systematically underestimate uncertainty in the magnitude range most relevant for regulatory decision-making.

3.3 The Poisson Process Model

Earthquake occurrence is usually modelled as a Poisson process, which assumes that events occur independently of one another at a constant average rate. Under this assumption, the number of earthquakes N occurring in a time interval t follows a Poisson distribution with rate parameter $\mu = \lambda t$:

$$P(N = k \mid \mu) = \frac{\mu^k e^{-\mu}}{k!} \tag{3.24}$$

where λ is the mean rate (events per unit time) and $k \in \{0, 1, 2, ...\}$ is the observed count.

The Poisson distribution has the key property that both its mean and variance equal μ:

$$E[N] = \mathrm{Var}(N) = \mu = \lambda t \tag{3.25}$$

However, there is evidence that this equality between the mean and variance does not hold for earthquake occurrence, with the variance being greater than the mean. This suggests

that the Poisson process may not be the most appropriate model for earthquake occurrence, and has prompted the use of alternative models, such as the negative binomial distribution (Gerstenberger et al., 2023). This issue is studied in detail later in the report.

For seismic hazard analysis, we are interested in the rate of events within specific magnitude ranges. From Equation 3.8, the rate of events with magnitude in the interval $[m_a, m_b)$ is:

$$\lambda(m_a \leq M < m_b) = \lambda_n \cdot [F_M(m_b) - F_M(m_a)] \tag{3.26}$$

This rate, multiplied by the observation time t, gives the expected number of events in that magnitude range over the observation period.

3.3.1 Likelihood for Event-Level Data

Given a catalogue of N earthquakes with observed magnitudes $\{m_1, m_2, ..., m_N\}$ and a total observation time T, the likelihood of the data under the GR model has two components:

1. **The number of events:** The probability of observing exactly N events follows the Poisson distribution with rate $\lambda_n T$:

$$P(N \mid \lambda_n, T) = \frac{(\lambda_n T)^N e^{-\lambda_n T}}{N!} \tag{3.27}$$

2. **The magnitude distribution:** Given that N events occurred, their magnitudes are independent draws from the GR distribution. The joint probability is:

$$P(m_1, ..., m_N \mid \beta, m_n, m_x) = \prod_{i=1}^{N} f_M(m_i \mid \beta, m_n, m_x) \tag{3.28}$$

The complete likelihood is the product of these two terms:

$$\mathcal{L}(\lambda_n, \beta \mid \{m_i\}, N, T) = \frac{(\lambda_n T)^N e^{-\lambda_n T}}{N!} \prod_{i=1}^{N} f_M(m_i \mid \beta) \tag{3.29}$$

Taking logarithms and dropping terms that do not depend on the parameters:

$$\ln \mathcal{L} \propto N \ln(\lambda_n) - \lambda_n T + \sum_{i=1}^{N} \ln f_M(m_i \mid \beta) \tag{3.30}$$

This event-level formulation is conceptually straightforward but difficult to implement in practice due to the need to account for catalogue completeness: older events may only be recorded above higher magnitude thresholds. So, we cannot use a constant T and must make extensions for variable completeness (Weichert, 1980).

3.4 From Continuous Magnitudes to Discrete Counts

The Poisson distribution operates on discrete counts, yet earthquake magnitudes are continuous variables. This creates a practical challenge: how do we construct a likelihood function that: (i) respects the discrete nature of the Poisson model; (ii) accommodates

the continuous magnitude distribution; and, (iii) handles variable completeness across the catalogue?

The solution is to discretise the magnitude range into bins. Let the magnitude range $[m_n, m_x]$ be partitioned into I bins, where bin i covers the interval $[m_{i-1}, m_i)$ with width $\Delta m_i = m_i - m_{i-1}$. For uniform bins (common in practice), $\Delta m_i = \Delta m$ for all i.

For each bin i, define:

- n_i: the observed number of events with magnitude in bin i
- t_i: the period of complete observation for bin i (years)
- $\mu_i = \lambda_i t_i$: the expected count, where λ_i is the rate of events in bin i

The rate of events in bin i is obtained from the GR distribution:

$$\lambda_i = \lambda_n [F_M(m_i) - F_M(m_{i-1})] = \lambda_n \gamma_i(\beta) \tag{3.31}$$

where we define the bin probability $\gamma_i(\beta)$ as:

$$\gamma_i(\beta) = \frac{e^{-\beta(m_{i-1}-m_n)} - e^{-\beta(m_i-m_n)}}{1 - e^{-\beta(m_x-m_n)}} \tag{3.32}$$

The γ_i terms represent the probability that an event falls in bin i, given that it has magnitude in $[m_n, m_x]$. By construction, $\sum_i \gamma_i(\beta) = 1$. The activity rate λ_n scales all bin rates proportionally, while β controls the shape – higher β concentrates probability in lower magnitude bins.

3.4.1 Binned Poisson Likelihood

Under the assumption that counts in different bins are independent Poisson random variables, the likelihood becomes:

$$\mathcal{L}(\lambda_n, \beta \mid \{n_i\}) = \prod_{i=1}^{I} \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!} \tag{3.33}$$

where $\mu_i = \lambda_n \gamma_i(\beta) t_i$.

The log-likelihood is:

$$\ln \mathcal{L} = \sum_{i=1}^{I} [n_i \ln(\mu_i) - \mu_i - \ln(n_i!)] \tag{3.34}$$

Substituting $\mu_i = \lambda_n \gamma_i t_i$ and dropping constants:

$$\ln \mathcal{L} = n_T \ln(\lambda_n) + \sum_{i=1}^{I} n_i \ln \gamma_i(\beta) - \lambda_n \sum_{i=1}^{I} \gamma_i(\beta) t_i \tag{3.35}$$

where $n_T = \sum_i n_i$ is the total number of observed events.

This is the form used by Johnston et al. (1994) and implemented in the PMLM approach described in Section 4.

3.5 Connection to Poisson Regression

The binned likelihood in Equation 3.34 has a natural interpretation as a Poisson regression (generalised linear model with log link and Poisson error distribution). This perspective provides both computational advantages and insight into the estimation problem.

In standard Poisson regression, we model the expected count in observation i as:

$$\ln(\mu_i) = \boldsymbol{x}_i^T \boldsymbol{\theta} + \text{offset}_i \tag{3.36}$$

where $\boldsymbol{x}_i$ is a vector of covariates, $\boldsymbol{\theta}$ is the parameter vector, and the offset accounts for known exposure differences (here, observation time t_i, or potentially areas as well depending upon what units are adopted for the rates or rate densities).

For the GR distribution, the expected count in bin i is:

$$\mu_i = \lambda_n \gamma_i(\beta) t_i \tag{3.37}$$

Taking logarithms:

$$\ln(\mu_i) = \ln(\lambda_n) + \ln(\gamma_i(\beta)) + \ln(t_i) \tag{3.38}$$

Comparing with Equation 3.36, we see that the GR model *almost* fits the Poisson regression framework, with $\ln(t_i)$ as an offset. However, there is a critical complication: the $\ln(\gamma_i(\beta))$ term is not linear in β.

Specifically, the bin probability involves the doubly-bounded exponential normalization:

$$\gamma_i(\beta) = \frac{e^{-\beta m_{i-1}} - e^{-\beta m_i}}{e^{-\beta m_n} - e^{-\beta m_x}} \tag{3.39}$$

The denominator $\left(e^{-\beta m_n} - e^{-\beta m_x}\right)$ accounts for the truncation at m_n and m_x, and makes γ_i a non-linear function of β. If we were to ignore this normalization (equivalent to assuming $m_x \to \infty$ and treating the distribution as singly-bounded), we would obtain:

$$\ln(\gamma_i) \approx -\beta(m_i - m_n) + \ln\left(1 - e^{-\beta \Delta m}\right) \tag{3.40}$$

which is approximately linear in β for small Δm.

This suggests an approximate approach: fit a standard Poisson regression with magnitude (or bin centre) as a linear predictor, then correct for the truncation. The exact approach, implemented in the PMLM and Bayesian methods described in this report, retains the full non-linear form and uses iterative optimization or MCMC to find the parameter estimates.

3.5.1 Practical Implications

The connection to Poisson regression has several practical implications:

1. **Software availability:** Standard GLM software can provide initial estimates, which can then be refined using the exact likelihood. This is particularly useful for diagnostic checks and model validation.

2. **Interpretation:** The GLM perspective clarifies that we are estimating a log-linear relationship between magnitude and occurrence rate, with β playing the role of a regression coefficient.
3. **Extensions:** Poisson regression naturally accommodates additional covariates (spatial variation, temporal trends, *etc.*) through the linear predictor, providing a framework for more complex models.
4. **Normalization:** The key distinction from standard GLM is the normalization constraint imposed by m_n and m_x. Ignoring this constraint leads to biased estimates, particularly when the magnitude range is narrow relative to the distribution's natural decay rate.

4 Current UK Practice

The penalised maximum likelihood method (PMLM) has traditionally been used to estimate the Gutenberg-Richter parameters λ_n and β from earthquake catalogues. The method is described in detail in Johnston et al. (1994), but this report is not readily available. Some details are provided in an appendix of Baker et al. (2021), but there is no coverage of how logic tree branches are obtained. Note that the original Johnston et al. (1994) report does not include any discussion of how one should represent epistemic uncertainties within a logic tree framework, but it does specify how to estimate the covariance matrix of the model parameters.

4.1 PMLM Implementation Details

Define α_n as the activity rate density above m_n, such that for a region of area A the total activity rate is $\lambda_n = \alpha_n \cdot A$ (the total number of events per year with magnitude at least as large as m_n). In the original presentation of Johnston et al. (1994), an overall region can be partitioned into a number of subregions, each with its own activity rate density $\alpha_{n,i}$ and area A_i. These sub-regions do not necessarily correspond to a seismic source zone, and zones can be divided into multiple sub-regions in order to reflect varying levels of catalogue completeness. This approach is relevant in the UK for zones associated with the North Sea where seismotectonic arguments might suggest uniform seismicity, but instrumental locations dictate that catalogue completeness is not uniform throughout the zones.

Although magnitude is fundamentally a continuous variable, and the GR distribution operates on continuous variables, the Johnston et al. (1994) method works by partitioning the range $[m_n, m_x]$ into discrete bins. Specifically, the magnitude range $[m_n, m_x]$ is partitioned into I bins (from 1 to I) such that $m_n = m_0$ and $m_x = m_I$. The width of these bins is not necessarily constant, but in many practical applications it will be (often using $\Delta m = 0.1$). Issues related to the effects of binning have been addressed in the literature (Bender, 1983).

For a given set of parameters λ_n and β, the rate of events in bin i (with magnitude range $[m_{i-1}, m_i)$) per year is:

$$\begin{aligned} \lambda_i &= \lambda_n [F_M(m_i \mid \beta, m_n, m_x) - F_M(m_{i-1} \mid \beta, m_n, m_x)] \\ &= \lambda_n \left[\frac{e^{-\beta(m_{i-1}-m_n)} - e^{-\beta(m_i-m_n)}}{1 - e^{-\beta(m_x-m_n)}} \right] \\ &= \lambda_n \left[\frac{e^{-\beta m_{i-1}} - e^{-\beta m_i}}{e^{-\beta m_n} - e^{-\beta m_x}} \right] \end{aligned} \tag{4.1}$$

Clearly, the activity rate in bin i scales linearly with λ_n. It is therefore convenient to express λ_i as: $\lambda_i = \lambda_n \gamma_i(\beta)$, where:

$$\gamma_i(\beta) = \frac{e^{-\beta m_{i-1}} - e^{-\beta m_i}}{e^{-\beta m_n} - e^{-\beta m_x}} \tag{4.2}$$

Because Equation 4.2 is defining the difference between the cdf evaluated at the upper and lower bounds of bin i, it is clear that it defines the probability of an event being in bin i and that:

$$\sum_{i=1}^{I} \gamma_i(\beta) = 1 \tag{4.3}$$

The parameter β just controls the shape of the distribution, and λ_n controls the overall rate of events. When the β value is relatively high, the probabilities in the lower i bins are reltively high, and *vice versa*

Equivalently, we could define the bin width $\Delta m_i = m_i - m_{i-1}$, and express $\gamma_i(\beta)$ as:

$$\gamma_i(\beta) = e^{-\beta(m_{i-1} - m_n)} \left[\frac{1 - e^{-\beta \Delta m_i}}{1 - e^{-\beta(m_x - m_n)}} \right] \tag{4.4}$$

Note that if we use uniform bin spacing (which is a very common situation), we then have:

$$\begin{aligned} \gamma_i(\beta) &= e^{-\beta(m_{i-1} - m_n)} \left[\frac{1 - e^{-\beta \Delta m}}{1 - e^{-\beta(m_x - m_n)}} \right] \\ &= e^{-\beta \Delta m (i-1)} \left[\frac{1 - e^{-\beta \Delta m}}{1 - e^{-\beta(m_x - m_n)}} \right] \end{aligned} \tag{4.5}$$

The term in the square brackets is then constant for all bins, and the $\gamma_i(\beta)$ terms form a geometric sequence with ratio $e^{-\beta \Delta m}$. Clearly, and importantly for later discussions, as β increases the $\gamma_i(\beta)$ terms decrease more rapidly with increasing bin index i.

This separation above allows us to clearly see the dependence of λ_i on both λ_n and β. The λ_n scales the entire distribution up or down, while β controls the relative probabilities of events in different magnitude bins.

Assuming a Poisson process, the probability of observing k_i events in the interval $[m_{i-1}, m_i)$ is:

$$P(K = k_i \mid \lambda_i, t_i) = \frac{(\lambda_i t_i)^{k_i} e^{-\lambda_i t_i}}{k_i!} \tag{4.6}$$

where t_i is the period of *complete* observation (in years) for bin i. Here, by complete we mean that all events with magnitude at least as large as m_{i-1} are included in the catalogue.

The likelihood of observing a particular earthquake catalogue with k_i events in each bin i is then:

$$\mathcal{L}(\lambda_n, \beta \mid \{k_i\}, \{t_i\}) = \prod_{i=1}^{I} P(K = k_i \mid \lambda_i, t_i) \tag{4.7}$$

Johnston et al. (1994) make some extensions for the case where completeness levels are not constant throughout a given region, but this change is not central to what follows. All we have to do to extend the method is to introduce the area of a sub-region A_j and define the rate of events in bin i for sub-region j as $\lambda_{ij} = \alpha_n A_j \gamma_i(\beta) = \lambda_{nj}\gamma_i(\beta)$, and then ensure we work with the appropriate t_{ij} and k_{ij} for the sub-region. These terms are then just used in:

$$P\big(K = k_{ij} \mid \lambda_{ij}, t_{ij}\big) = \frac{\big(\lambda_{ij} t_{ij}\big)^{k_{ij}} e^{-\lambda_{ij} t_{ij}}}{k_{ij}!} \tag{4.8}$$

and the likelihood function becomes:

$$\mathcal{L}\big(\lambda_n, \beta \mid \{k_{ij}\}, \{t_{ij}\}\big) = \prod_{j=1}^{J} \prod_{i=1}^{I} P\big(K = k_{ij} \mid \lambda_{ij}, t_{ij}\big) \tag{4.9}$$

where J is the number of sub-regions.

4.1.1 The Penalised Maximum Likelihood Method

Following Dyck (1985); Veneziano & Dyck, 1985, the above likelihood function is modified to introduce a prior on β (equivalent to a prior on b) such that the likelihood function becomes:

$$\mathcal{L}(\lambda_n, \beta \mid \{k_i\}, \{t_i\}) = \left(\prod_{i=1}^{I} P(K = k_i \mid \lambda_i, t_i)\right) \exp\left(-\frac{\big(\beta - \beta_p\big)^2}{2\sigma_{\beta,p}^2}\right) \tag{4.10}$$

Note that their notation uses a penalty W such that $W = 1/\big(\sigma_{\beta,p}^2\big)$. Explicitly, they write:

$$\mathcal{L}(\lambda_n, \beta \mid \{k_i\}, \{t_i\}) = \left(\prod_{i=1}^{I} P(K = k_i \mid \lambda_i, t_i)\right) \exp\left(-\frac{W}{2}\big(\beta - \beta_p\big)^2\right) \tag{4.11}$$

This *weight* W is poorly understood within the community, and it is not clear why Johnston et al. (1994) decided to frame it as a weight rather than be more explicit. A common value adopted in UK practice is $W = 25$, and this is often attributed to Musson (2011), but while that report makes use of this value, it does not appear to justify its choice. In fact, the report actually states: *"It is suggested by Johnston et al. (1994) that the weight be inverse to the variance of the prior value; in practice it is a subjective decision."* Regardless, the value of $W = 25$ is often adopted, and given the understanding of its actual meaning, it is clear that it corresponds to a prior on the β value of $\sigma_{\beta,p} = 1/\sqrt{25} = 0.2$, and that as $\beta = \log(10) \cdot b$, this corresponds to a prior on b of $\sigma_{b,p} = 0.2/\ln(10) \approx 0.087$. So, the conventional practice in the UK is to use a Gaussian prior on b with mean 1.0 and standard deviation 0.087.

4.1.2 Equivalence with Maximum A Posteriori

The above section explains where the name *penalised maximim likelihood method* comes from. However, if we recall that in a Bayesian context the posterior distribution, $\pi(\boldsymbol{\theta} \mid \boldsymbol{x})$, is proportional to the likelihood, $\mathcal{L}(\boldsymbol{x} \mid \boldsymbol{\theta})$, times the prior, $\pi(\boldsymbol{\theta})$:

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{x}) \propto \mathcal{L}(\boldsymbol{x} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{4.12}$$

If we find that parameters that correspond to the maximum of the posterior distribution, those are referred to as the maximum a posteriori (MAP) estimate. So, if we use a naive or flat prior for the activity rate, but impose a Gaussian prior on β, then the MAP estimate is equivalent to the penalised maximum likelihood estimate.

4.1.3 Evaluation of the Penalised Maximum Likelihood

Working with the case where we have single area (rather than sub-regions), the log-likelihood is then:

$$\begin{aligned} \ln \mathcal{L}(\lambda_n, \beta \mid \{k_i\}, \{t_i\}) &= \sum_{i=1}^{I} [k_i \ln(\lambda_i t_i) - \lambda_i t_i - \ln(k_i!)] - \frac{\left(\beta - \beta_p\right)^2}{2\sigma_{\beta,p}^2} \\ &= \sum_{i=1}^{I} [k_i \ln(\lambda_i) + k_i \ln(t_i) - \lambda_i t_i - \ln(k_i!)] - \frac{\left(\beta - \beta_p\right)^2}{2\sigma_{\beta,p}^2} \end{aligned} \tag{4.13}$$

For a given catalogue, the k_i and t_i are fixed, so the term $\sum_{i=1}^{I} k_i \ln(t_i)$ is constant with respect to the parameters λ_n and β, and the factorial terms also do not depend on the parameters. Therefore, when maximising the log-likelihood we can ignore these constant terms, giving:

$$\ln \mathcal{L}(\lambda_n, \beta \mid \{k_i\}, \{t_i\}) \propto \sum_{i=1}^{I} [k_i \ln(\lambda_i) - \lambda_i t_i] - \frac{\left(\beta - \beta_p\right)^2}{2\sigma_{\beta,p}^2} \tag{4.14}$$

The derivative of the log-likelihood with respect to λ_n and β can then be used to find the maximum likelihood estimates of these parameters.

4.1.4 Partial Derivatives of the Penalised Log-Likelihood

The λ_i terms are functions of both λ_n and β, and we have $\ln(\lambda_i)$ and λ_i in the log-likelihood expression. However, we know that in general:

$$\frac{\partial \ln(\lambda_i)}{\partial x} = \left(\frac{1}{\lambda_i}\right)\frac{\partial \lambda_i}{\partial x} \quad \Leftrightarrow \quad \frac{\partial \lambda_i}{\partial x} = \lambda_i \frac{\partial \ln(\lambda_i)}{\partial x} \tag{4.15}$$

for any variable x.

So, the partial derivatives of the log-likelihood with respect to λ_n and β are:

$$\frac{\partial \ln \mathcal{L}}{\partial \lambda_n} = \sum_{i=1}^{I} \left[k_i \left(\frac{1}{\lambda_i}\right) \frac{\partial \lambda_i}{\partial \lambda_n} - t_i \frac{\partial \lambda_i}{\partial \lambda_n} \right] \tag{4.16}$$

and

$$\frac{\partial \ln \mathcal{L}}{\partial \beta} = \sum_{i=1}^{I}\left[k_i\left(\frac{1}{\lambda_i}\right)\frac{\partial \lambda_i}{\partial \beta} - t_i\frac{\partial \lambda_i}{\partial \beta}\right] - \left(\frac{\beta - \beta_p}{\sigma^2_{\beta,p}}\right) \tag{4.17}$$

Recall that λ_i is a function of both λ_n and β as defined above. The partial derivative of λ_i with respect to λ_n is simply:

$$\frac{\partial \lambda_i}{\partial \lambda_n} = \gamma_i(\beta) \tag{4.18}$$

This means that our expression for the partial derivative of the log-likelihood with respect to λ_n becomes:

$$\begin{aligned}\frac{\partial \ln \mathcal{L}}{\partial \lambda_n} &= \sum_{i=1}^{I}\left[k_i\left(\frac{1}{\lambda_i}\right)\gamma_i(\beta) - t_i\gamma_i(\beta)\right] \\ &= \sum_{i=1}^{I}\left[\frac{k_i}{\lambda_n} - t_i\frac{\lambda_i}{\lambda_n}\right] \\ &= \frac{1}{\lambda_n}\sum_{i=1}^{I}[k_i - t_i\lambda_i]\end{aligned} \tag{4.19}$$

For the partial derivative with respect to β it is more convenient to work in terms of $\frac{\partial \ln(\lambda_i)}{\partial \beta}$ than $\frac{\partial \lambda_i}{\partial \beta}$ directly. This is because, from Equation 4.15, we have the $\frac{1}{\lambda_i}$ term that allows for some cancellations to arise.

For simplicity, let:

$$D_i(\beta) = \frac{\partial \ln(\lambda_i)}{\partial \beta} = \left[\frac{m_i e^{-\beta m_i} - m_{i-1}e^{-\beta m_{i-1}}}{e^{-\beta m_{i-1}} - e^{-\beta m_i}} - \frac{m_x e^{-\beta m_x} - m_n e^{-\beta m_n}}{e^{-\beta m_n} - e^{-\beta m_x}}\right] \tag{4.20}$$

If we let $\Delta m = m_i - m_{i-1}$ again for constant bin widths, we have:

$$D_i(\beta) = -(i-1)\Delta m + \left[\frac{\Delta m}{e^{\beta \Delta m} - 1} - \frac{m_x - m_n}{e^{\beta(m_x - m_n)} - 1}\right] \tag{4.21}$$

This gives us the partial derivative of the log-likelihood with respect to β as:

$$\begin{aligned}\frac{\partial \ln \mathcal{L}}{\partial \beta} &= \sum_{i=1}^{I}[k_i D_i(\beta) - t_i\lambda_i D_i(\beta)] - \left(\frac{\beta - \beta_p}{\sigma^2_{\beta,p}}\right) \\ &= \sum_{i=1}^{I}[(k_i - t_i\lambda_i)D_i(\beta)] - \left(\frac{\beta - \beta_p}{\sigma^2_{\beta,p}}\right)\end{aligned} \tag{4.22}$$

Although we have the two parameters λ_n and β that we're interested in estimating, we know that λ_i depends linearly on λ_n. So, β controls the shape of the magnitude distribution, while λ_n controls the overall scaling of the activity rate. Both of the expressions for the partial derivatives of the log-likelihood contain the term $(k_i - t_i\lambda_i)$, which represents the difference between the observed number of events in bin i and the expected number of events in bin i given the model parameters (we will later call this the residual). While the prior on

β complicates things a little, we can see that when the $(k_i - t_i\lambda_i) \to 0$ the overall partial derivatives will also tend to zero. As the $\{k_i\}$ and $\{t_i\}$ terms are fixed for a given catalogue, and we know that $\lambda_i = \lambda_n\gamma_i(\beta)$, we can see that the partial derivatives can match a particular value for different combinations of λ_n and β. Specifically, if we increase λ_n we must decrease $\gamma_i(\beta)$. if we increase β (steepen the magnitude distribution), we can compensate for this by increasing λ_n (increasing the overall activity rate). This trade-off between β and λ_n is illustrated in Figure 4.1.

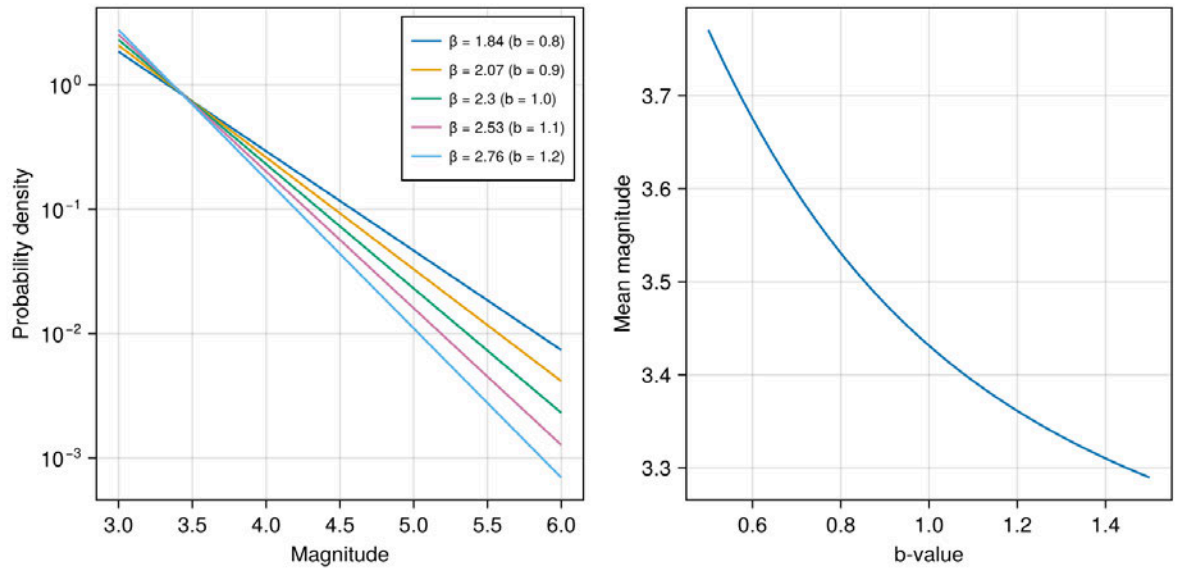

Figure 4.1: The left panel shows the probability density function, $f_M(m)$, for the GR distribution for $m_n = 3.0$ and $m_x = 6.0$ for a series of b-values shown in the legend. The right panel shows how the mean magnitude ($\int_{m_n}^{m_x} m f_M(m) dm$) of the distribution varies with the value of b-value.

The left panel of Figure 4.1 shows that when we move to lower values of β, the distribution becomes shallower, or flatter, and the mean magnitude of the distribution increases (shown in the right panel). The mean magnitude is defined as:

$$\mathbb{E}[m] = \int_{m_n}^{m_x} m f_M(m) dm = m_n + \frac{1}{\beta} - \frac{(m_x - m_n)e^{-\beta(m_x - m_n)}}{1 - e^{-\beta(m_x - m_n)}} \tag{4.23}$$

Because the $(k_i - t_i\lambda_i)$ terms must be evaluated over all magnitude bins, and that we will have sampling variability in the counts of events in each bin, there will be combinations of λ_n and β that give the same likelihood value (with those combinations performing better or worse in any given magnitude bin).

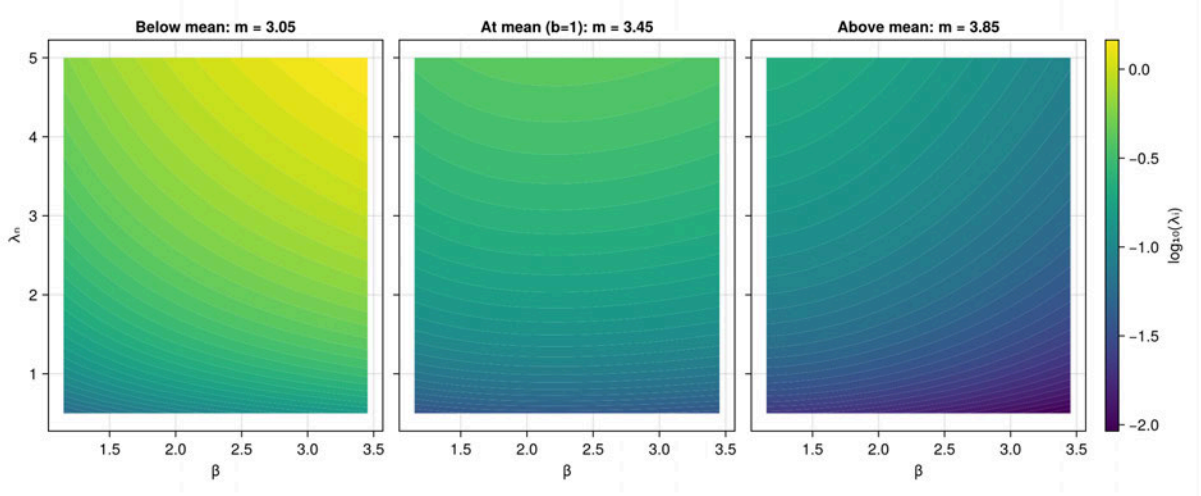

Figure 4.2: Plots of the λ_i as a function of β and λ_n for a selection of magnitude bins (as labelled in the legend). The contours show lines of constant λ_i for each bin, illustrating the trade-off between β and λ_n to achieve the same activity rate in each bin.

Figure 4.2 demonstrates this relationship by showing contours of constant λ_i against β and λ_n for a selection of magnitude bins. The magnitude bins have been chosen to represent cases where we are below, around or above the mean magnitude for the distribution. As seen previously in Figure 4.1, if we assume that $b = 1$ is some nominal value, we can see that when we increase that value (steepen the distribution) the mean magnitude reduces slightly, and rates of occurrence increase (relative to the $b = 1$ case) for the lower magnitude bins, while rates reduce for the higher magnitude bins. The opposite is true when we reduce the b-value (flatten the distribution). Figure 4.2 shows how we can compensate for these changes in β by adjusting λ_n to achieve the same activity rate in each bin. The left panel shows that rates tend to increase as we move in the $\{\beta, \lambda_n\}$ space from lower left to upper right, while the right panel shows the opposite trend. The central panel, which corresponds to a magnitude bin around the mean magnitude, shows that the activity rate is primarily controlled by movements up and down (the rate is driven by λ_n and is relatively insensitive to β).

Now, if we return to the expressions for the partial derivatives of the log-likelihood with respect to the parameters, and start with that for λ_n:

$$\frac{\partial \ln \mathcal{L}}{\partial \lambda_n} = \frac{1}{\lambda_n} \sum_{i=1}^{I} [k_i - t_i \lambda_i] \tag{4.24}$$

Set this equal to zero to find the maximum likelihood estimate:

$$0 = \frac{1}{\lambda_n} \sum_{i=1}^{I} [k_i - t_i \lambda_i] \tag{4.25}$$

and rearrange to give:

$$\sum_{i=1}^{I} k_i = \lambda_n \sum_{i=1}^{I} t_i \gamma_i(\beta) \tag{4.26}$$

Hence, the maximum likelihood estimate of λ_n for a given value of β is:

$$\lambda_n^* = \frac{\sum_{i=1}^{I} k_i}{\sum_{i=1}^{I} t_i \gamma_i(\beta)} = \frac{N}{\sum_{i=1}^{I} t_i \gamma_i(\beta)} \tag{4.27}$$

with $N = \sum_{i=1}^{I} k_i$ being the total count of events over all bins. Recall that $\gamma_i(\beta)$ is simply the probability that an event has a magnitude in bin i for a given value of β. So, the denominator here is effectively the weighted average of the periods of observation for the catalogue. The crucial aspect of this equation is that the N is fixed by the observed catalogue, and so λ_n and β must trade-off with each other. As β increases, the probabilities of events in relatively large magntiude bins (associated with long t_i values) decreases. Therefore, as β increases, the denominator tends to decrease and λ_n increases. This is the origin of the positive correlation that exists between β and λ_n.

Then, return to the partial derivative with respect to β:

$$\frac{\partial \ln \mathcal{L}}{\partial \beta} = \sum_{i=1}^{I} [(k_i - t_i \lambda_i) D_i(\beta)] - \left(\frac{\beta - \beta_p}{\sigma_{\beta,p}^2} \right) \tag{4.28}$$

and substitute in $\lambda_i = \lambda_n \gamma_i(\beta)$

$$\frac{\partial \ln \mathcal{L}}{\partial \beta} = \sum_{i=1}^{I} [(k_i - t_i \lambda_n \gamma_i(\beta)) D_i(\beta)] - \left(\frac{\beta - \beta_p}{\sigma_{\beta,p}^2} \right) \tag{4.29}$$

Grouping terms is instructive:

$$\frac{\partial \ln \mathcal{L}}{\partial \beta} = \underbrace{\sum_{i=1}^{I} k_i D_i(\beta)}_{\substack{\text{Data-weighted} \\ \text{Sensitivity}}} - \lambda_n \underbrace{\sum_{i=1}^{I} t_i \gamma_i(\beta) D_i(\beta)}_{\substack{\text{Exposure-weighted} \\ \text{Sensitivity}}} - \underbrace{\left(\frac{\beta - \beta_p}{\sigma_{\beta,p}^2} \right)}_{\text{Regularisation}} \tag{4.30}$$

In this representation, we see that λ_n is nicely isolated and given that we already have our expression for λ_n corresponding to $\frac{\partial \ln \mathcal{L}}{\partial \lambda_n} = 0$ we can substitute this solution to obtain an expression entirely in terms of β:

$$\frac{\partial \ln \mathcal{L}}{\partial \beta} = \sum_{i=1}^{I} k_i D_i(\beta) - N \frac{\sum_{i=1}^{I} t_i \gamma_i(\beta) D_i(\beta)}{\sum_{i=1}^{I} t_i \gamma_i(\beta)} - \left(\frac{\beta - \beta_p}{\sigma_{\beta,p}^2} \right) \tag{4.31}$$

And, we want to find the β that satisfies:

$$0 = \sum_{i=1}^{I} k_i D_i(\beta) - N \frac{\sum_{i=1}^{I} t_i \gamma_i(\beta) D_i(\beta)}{\sum_{i=1}^{I} t_i \gamma_i(\beta)} - \left(\frac{\beta - \beta_p}{\sigma_{\beta,p}^2} \right) \tag{4.32}$$

Dividing all terms by N is equivalent to writing:

$$0 = \frac{\sum_{i=1}^{I} k_i D_i(\beta)}{\sum_{i=1}^{I} k_i} - \frac{\sum_{i=1}^{I} t_i \gamma_i(\beta) D_i(\beta)}{\sum_{i=1}^{I} t_i \gamma_i(\beta)} - \frac{1}{N}\left(\frac{\beta - \beta_p}{\sigma_{\beta,p}^2}\right) \tag{4.33}$$

This representation is useful for a couple of reasons: (1) it clearly shows that for large N the regularisation term related to the prior becomes less important (as one would expect); and, (2) the optimal β value can be seen to correspond to balancing the data-weighted and exposure-weighted sensitivities (as the first two terms are weighted averages of the same $D_i(\beta)$ expressions).

Ultimately, we see that we can solve for the optimal β and λ by using the above expression to first find β^* and know that the corresponding value of λ_n^* immediately arises as a result. That is, rather than solving a two-parameter optimisation for $\{\beta, \lambda_n\}$, we really solve a profiled optimisation for β and then obtain the matched λ_n value.

In the limit where $N \to \infty$ the regularisation term disappears, and we just have to satisfy:

$$\frac{\sum_{i=1}^{I} k_i D_i(\beta)}{\sum_{i=1}^{I} k_i} = \frac{\sum_{i=1}^{I} t_i \gamma_i(\beta) D_i(\beta)}{\sum_{i=1}^{I} t_i \gamma_i(\beta)} \tag{4.34}$$

which we can appreciate involves matching k_i to $t_i\gamma_i(\beta)$ as closely as possible, on average, over all bins.

4.1.5 Parameter Uncertainties

Once the optimal parameters are identified, the covariance of these parameters can be computed. The standard procedure makes use of the Hessian matrix:

$$\boldsymbol{H} = \begin{bmatrix} \frac{\partial^2 \ln \mathcal{L}}{\partial \lambda_n^2} & \frac{\partial^2 \ln \mathcal{L}}{\partial \lambda_n \partial \beta} \\ \frac{\partial^2 \ln \mathcal{L}}{\partial \lambda_n \partial \beta} & \frac{\partial^2 \ln \mathcal{L}}{\partial \beta^2} \end{bmatrix} = \begin{bmatrix} H_{\lambda\lambda} & H_{\lambda\beta} \\ H_{\lambda\beta} & H_{\beta\beta} \end{bmatrix} = -\boldsymbol{\mathcal{I}} \tag{4.35}$$

The Fisher information matrix is the negative Hessian, $\boldsymbol{\mathcal{I}} = -\boldsymbol{H}$, as shown above, and the covariance matrix of the model parameters is defined as the inverse of this information matrix: $\boldsymbol{\Sigma} = \boldsymbol{\mathcal{I}}^{-1}$.

We can define the terms of the Fisher information matrix as:

$$\boldsymbol{\mathcal{I}} = \begin{bmatrix} A & B \\ B & C \end{bmatrix} = \begin{bmatrix} -H_{\lambda\lambda} & -H_{\lambda\beta} \\ -H_{\lambda\beta} & -H_{\beta\beta} \end{bmatrix} \tag{4.36}$$

Because we have maximimised the log-likelihood function (and we assume that it is continuous and twice differentiable), the curvature at the optimal parameters is negative. When we invert the Fisher information matrix, we are relating the variances of the parameters to the inverse of the curvature at the optimal parameters. High curvature maps to small variances, and well-constrained solutions, and *vice versa*.

The terms of the information matrix are defined by:

$$A = -H_{\lambda\lambda} = \frac{\sum_{i=1}^{I} k_i}{\lambda_n^2} = \frac{N}{\lambda_n^2} \tag{4.37}$$

which is the curvature associated with the activity rate, and N is the total count of events in the catalogue. The curvature associated with β is:

$$C = -H_{\beta\beta} \approx \sum_{i=1}^{I} t_i \lambda_i [D_i(\beta)]^2 + \frac{1}{\sigma_{\beta,p}^2} \tag{4.38}$$

and the interaction term is:

$$B = -H_{\lambda\beta} = \sum_{i=1}^{I} t_i \gamma_i(\beta) D_i(\beta) \tag{4.39}$$

Because the covariance matrix is the inverse of the Fisher information matrix, we have:

$$\boldsymbol{\Sigma} = \boldsymbol{\mathcal{I}}^{-1} = \frac{1}{AC - B^2} \begin{bmatrix} C & -B \\ -B & A \end{bmatrix} = \begin{bmatrix} \sigma_{\lambda_n}^2 & \mathrm{cov}(\lambda_n, \beta) \\ \mathrm{cov}(\lambda_n, \beta) & \sigma_\beta^2 \end{bmatrix} \tag{4.40}$$

where $\sigma_{\lambda_n}^2$ is the variance of the activity rate, σ_β^2 is the variance of the Gutenberg-Richter slope, and $\mathrm{cov}(\lambda_n, \beta)$ is the covariance between the activity rate and the Gutenberg-Richter slope. For convenience, define the determinant of the information matrix as:

$$\mathcal{D} = AC - B^2 \tag{4.41}$$

Making the relevant substitutions, the variance of the activity rate is:

$$\sigma_{\lambda_n}^2 = \frac{C}{\mathcal{D}} = \frac{\sum_{i=1}^{I} t_i \lambda_i D_i^2 + \frac{1}{\sigma_{\beta,p}^2}}{\left(\frac{N}{\lambda_n^2}\right)\left(\sum_{i=1}^{I} t_i \lambda_i D_i^2 + \frac{1}{\sigma_{\beta,p}^2}\right) - \left(\sum_{i=1}^{I} t_i \gamma_i D_i\right)^2} \tag{4.42}$$

which we can also write as:

$$\sigma_{\lambda_n}^2 = \frac{1}{A - \frac{B^2}{C}} = \frac{1}{\underbrace{\frac{N}{\lambda_n^2}}_{\substack{\text{Raw} \\ \text{curvature}}} - \underbrace{\frac{\left(\sum_{i=1}^{I} t_i \gamma_i D_i\right)^2}{\sum_{i=1}^{I} t_i \lambda_i D_i^2 + \frac{1}{\sigma_{\beta,p}^2}}}_{\text{Precision loss}}} \tag{4.43}$$

Similarly, the variance of the Gutenberg-Richter slope is:

$$\sigma_\beta^2 = \frac{A}{\mathcal{D}} = \frac{\frac{N}{\lambda_n^2}}{\left(\frac{N}{\lambda_n^2}\right)\left(\sum_{i=1}^{I} t_i \lambda_i D_i^2 + \frac{1}{\sigma_{\beta,p}^2}\right) - \left(\sum_{i=1}^{I} t_i \gamma_i D_i\right)^2} \tag{4.44}$$

which we can also write as:

$$\sigma_\beta^2 = \frac{1}{C - B^2/A} = \frac{1}{\underbrace{\sum_{i=1}^{I} t_i \lambda_i D_i^2 + \frac{1}{\sigma_{\beta,p}^2}}_{\text{Raw curvature}} - \underbrace{\left(\frac{\lambda_n^2}{N}\right)\left(\sum_{i=1}^{I} t_i \gamma_i D_i\right)^2}_{\text{Precision loss}}} \tag{4.45}$$

These alternative representations show that the effect of the interaction term (the correlation between the parameters) is to dilute the precision in the estimate of the individual parameters. For β, it also shows that the raw curvature depends upon the prior weighting assigned to the slope parameter. The stronger we make the prior (leading to smaller $\sigma_{\beta,p}$), the larger the raw curvature.

The covariance between the activity rate and the Gutenberg-Richter slope is:

$$\mathrm{cov}(\lambda_n, \beta) = -\frac{B}{\mathcal{D}} \tag{4.46}$$

and the correlation coefficient is:

$$\rho_{\lambda_n \beta} = \frac{\mathrm{cov}(\lambda_n, \beta)}{\sigma_{\lambda_n}\sigma_\beta} = -\frac{B}{\sqrt{CA}} \tag{4.47}$$

Making the relevant subsitutions, we have:

$$\rho_{\lambda_n \beta} = -\frac{\lambda_n}{\sqrt{N}} \frac{\sum_{i=1}^{I} t_i \gamma_i D_i}{\sqrt{\sum_{i=1}^{I} t_i \lambda_i D_i^2 + \frac{1}{\sigma^2_{\beta,p}}}} \tag{4.48}$$

Here, again, we see that the prior specification has an effect on the correlation between the parameters – with the stronger the prior, the lower the correlation. However, we also see that the magnitude of the correlation reduces as the total catalogue size N increases. In general, both factors tell the same story – the more precisely we can resolve the parameters (either through data constraints or prior specification), the lower the correlation between the parameters.

Figure 4.3 shows how the various terms that appear within the expressions for the variances or correlations scale with magnitude for a range of β values. By construction, the variance terms must be positive. For the correlation, we see that the numerator sums over $\gamma_i D_i$ contributions that are scaled by the exposure time t_i. This numerator controls the sign of the correlation because both terms on the denominator are positive. This can be appreciated from the bottom right panel of Figure 4.3 – which shows that $\gamma_i D_i^2 \geq 0$, but we know $\lambda_i = \lambda_n \gamma_i$ and $\lambda_n > 0$. The sign of the correlation is therefore controlled by the sign of $\sum_{i=1}^{I} t_i \gamma_i D_i$ and the bottom left panel of Figure 4.3 shows that $\gamma_i D_i > 0$ for $m < \mathbb{E}(m)$ and that $\gamma_i D_i < 0$ for $m > \mathbb{E}(m)$.

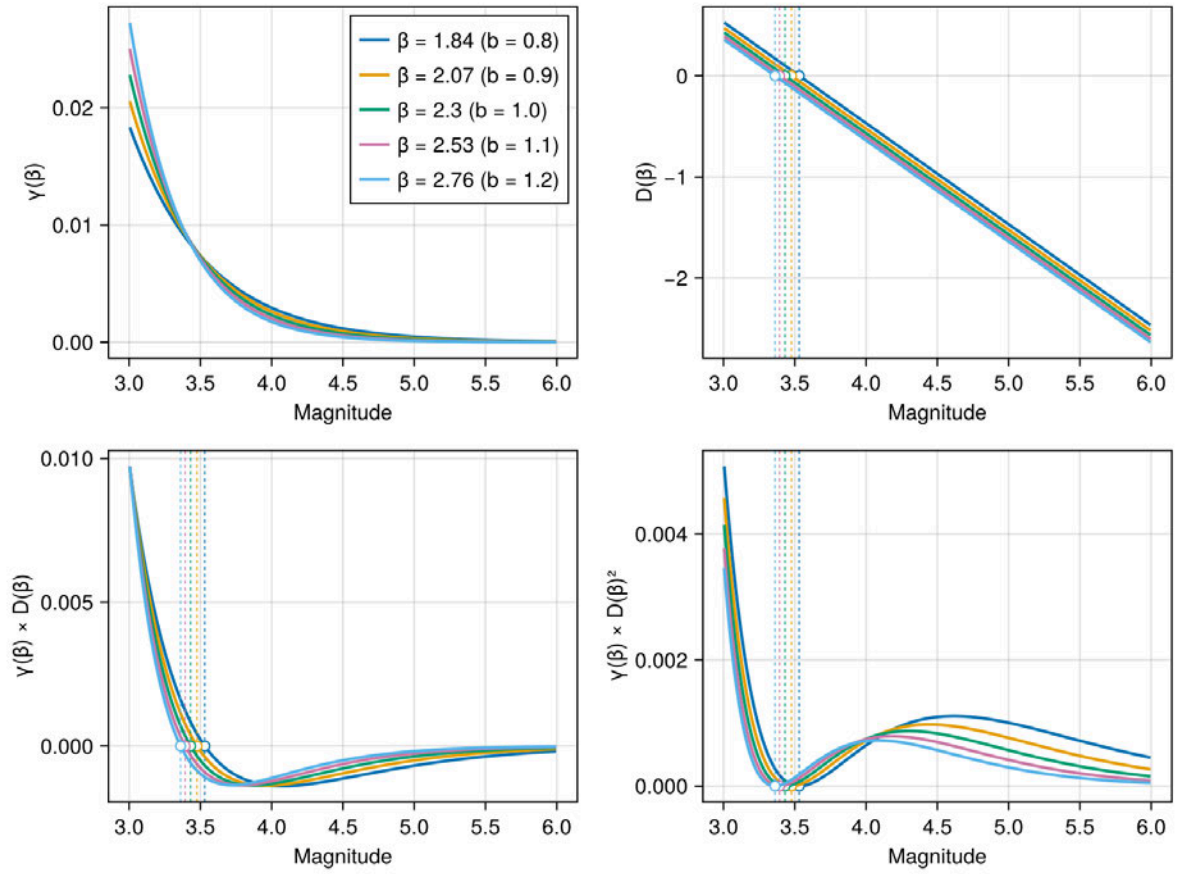

Figure 4.3: Scaling of the terms $\gamma_i(\beta)$ (top left), $D_i(\beta)$ (top right), $\gamma_i(\beta)D_i(\beta)$ (bottom left), and $\gamma_i(\beta)[D_i(\beta)]^2$ (bottom right), that appear in various expressions related to the covariance terms. Vertical dashed lines show the mean magnitude, $\mathbb{E}(m)$, for each distribution, and markers show $\{\mathbb{E}(m), 0\}$ in each case

Recall that $D_i(\beta) = \frac{\partial \ln(\lambda_i)}{\partial \beta}$, and that:

$$\frac{\partial \lambda_i}{\partial \beta} = \lambda_i \frac{\partial \ln(\lambda_i)}{\partial \beta} = \lambda_i D_i \tag{4.49}$$

and as $\lambda_i = \lambda_n \gamma_i(\beta)$, we have:

$$\frac{\partial \lambda_i}{\partial \beta} = \lambda_n \gamma_i D_i \quad \Rightarrow \quad \frac{\partial \gamma_i}{\partial \beta} = \gamma_i D_i \tag{4.50}$$

The numerator of the correlation equation can therefore be written as:

$$-\lambda_n \sum_{i=1}^{I} t_i \gamma_i D_i \equiv -\lambda_n \sum_{i=1} t_i \frac{\partial \gamma_i}{\partial \beta} \tag{4.51}$$

Consider, for a moment, the case where $t_i = T$ for all i. Then, the numerator of the correlation equation is:

$$\lambda_n \sum_{i=1} t_i \frac{\partial \gamma_i}{\partial \beta} = \lambda_n T \sum_{i=1} \frac{\partial \gamma_i}{\partial \beta} \tag{4.52}$$

and we know that the sum of the partial derivatives is the same as the partial derivative of the sum:

$$\sum_{i=1} \frac{\partial \gamma_i}{\partial \beta} = \frac{\partial}{\partial \beta}\left(\sum_{i=1} \gamma_i\right) \tag{4.53}$$

Now, γ_i is just the probability of an event falling in magnitude bin i, given the Gutenberg-Richter slope β. The summation over all magnitude bins is equal to 1, by definition, and we therefore have:

$$\frac{\partial}{\partial \beta}\underbrace{\left(\sum_{i=1} \gamma_i\right)}_{=1} = 0 \tag{4.54}$$

and so the numerator of the correlation equation is 0. So, for uniform observation periods $t_i = T$ for all i, the correlation between λ_n and β is 0.

In reality, however, the observation periods are not uniform, and there is a systematic trend where larger magnitudes have longer obserations periods than for samll magnitude events. This means that the negative $\gamma_i D_i$ combinations for magnitudes above the mean magnitude will have a greater impact on the correlation than the positive $\gamma_i D_i$ combinations for magnitudes below the mean magnitude. And, as the sign of the correlation is opposite to the sign of the sum: $\sum_{i=1}^{I} t_i \gamma_i D_i$, the correlation will be positive.

4.1.6 Additional details regarding $H_{\beta\beta}$

Note that earlier we approximated the curvature associated with β in Equation 4.38, but this approximation is not strictly required, it just simplifies the maths and is valid for cases where the residuals are small (with the residuals being $k_i - t_i\lambda_i$). The general structure for the partial derivative of the log-likelihood function with respect to β was:

$$\frac{\partial \ln \mathcal{L}}{\partial \beta} = \sum_{i=1}^{I} \underbrace{(k_i - t_i\lambda_i)}_{\text{Residual}} D_i(\beta) - \left(\frac{\beta - \beta_p}{\sigma^2_{\beta,p}}\right) \tag{4.55}$$

When we differentiate this expression with respect to β again, the regularisation term provides a simple $1/\sigma^2_{\beta,p}$ contribution, but the product term within the summation is more elaborate (recall that $\lambda_i = \lambda_n \gamma_i(\beta)$). Using the product rule $(uv)' = u'v + uv'$, we have two contributions:

- from differentiating $(k_i - t_i\lambda_i)$ we get $-t_i \frac{\partial \lambda_i}{\partial \beta}$ and from previous derivations we know that $\frac{\partial \lambda_i}{\partial \beta} = \lambda_i \frac{\partial \ln(\lambda_i)}{\partial \beta} \equiv \lambda_i D_i$. Hence, differentiating $(k_i - t_i\lambda_i)$ leads to contributions that scale with $[D_i(\beta)]^2$
- from differentiating $D_i(\beta)$ we obtain $\frac{\partial D_i(\beta)}{\partial \beta}$, but this is multiplied by the residuals $(k_i - t_i\lambda_i)$.

Because we are evaluating these expressions at our optimal parameters, we should expect that the residual terms are close to zero and that they will have a mean of zero. We can therefore drop the second contribution related to differentiating $D_i(\beta)$, and this is where the approximation in Equation 4.38 comes from.

However, for completeness, and because the numerical implementations will often work with the full expressions, we can define the exact expression as:

$$H_{\beta\beta} = \sum_{i=1}^{I} \left[\underbrace{-t_i \lambda_i [D_i(\beta)]^2}_{\text{Primary curvature}} + \underbrace{(k_i - t_i \lambda_i) S_i(\beta)}_{\text{Residual curvature}} \right] - \frac{1}{\sigma_{\beta,p}^2} \tag{4.56}$$

where

$$S_i(\beta) = \frac{\partial^2 \ln(\lambda_i)}{\partial \beta^2} = \frac{\partial D_i(\beta)}{\partial \beta} \tag{4.57}$$

Recall that:

$$D_i(\beta) = \left[\frac{m_i e^{-\beta m_i} - m_{i-1} e^{-\beta m_{i-1}}}{e^{-\beta m_{i-1}} - e^{-\beta m_i}} - \frac{m_x e^{-\beta m_x} - m_n e^{-\beta m_n}}{e^{-\beta m_n} - e^{-\beta m_x}} \right] \tag{4.58}$$

and represent this as:

$$D_i(\beta) = \frac{U'}{U} - \frac{V'}{V} \tag{4.59}$$

where

$$U = e^{-\beta m_{i-1}} - e^{-\beta m_i} \quad \Rightarrow \quad U' = m_i e^{-\beta m_i} - m_{i-1} e^{-\beta m_{i-1}} \tag{4.60}$$

and

$$V = e^{-\beta m_n} - e^{-\beta m_x} \quad \Rightarrow \quad V' = m_x e^{-\beta m_x} - m_n e^{-\beta m_n} \tag{4.61}$$

The full expression for $S_i(\beta)$ is then defined in terms of these components:

$$S_i(\beta) = \left(\frac{U'' U - (U')^2}{U^2} \right) - \left(\frac{V'' V - (V')^2}{V^2} \right) \tag{4.62}$$

The only terms not fully defined thus far are:

$$U'' = m_{i-1}^2 e^{-\beta m_{i-1}} - m_i^2 e^{-\beta m_i} \quad \text{and} \quad V'' = m_n^2 e^{-\beta m_n} - m_x^2 e^{-\beta m_x} \tag{4.63}$$

4.2 Logic tree representations

The PMLM method is commonly used in UK practice, and when logic trees are constructed to represent the parametric uncertainty it is common to cite Johnston et al. (1994) despite the fact that this document does not cover how uncertainties should be represented in logic trees. It is rare to see any equations specified for defining logic tree weights, but the general approach adopted in practice appears to be:

- Compute the covariance matrix for the parameters as shown in the previous sections
- Define a grid of $\{\lambda_n, \beta\}$ (or $\{\alpha_n, \beta\}$) combinations by taking $\lambda \pm \varepsilon\sigma_{\lambda_n}$ and $\beta \pm \varepsilon\sigma_\beta$ with $\varepsilon \in \{-1, 0, +1\}$ or $\varepsilon \in \{-2, -1, 0, +1, +2\}$. This results in a 3×3 or 5×5 grid of combinations
- Evaluate the penalised log-likelihood function at each of these grid points $\ln \mathcal{L}(\lambda_{n,i}, \beta_j)$, and define weights as:

$$w_{ij} = \frac{q_{i,j} \ln \mathcal{L}(\lambda_{n,i}, \beta_j)}{\sum_j \sum_i q_{i,j} \ln \mathcal{L}(\lambda_{n,i}, \beta_j)} \tag{4.64}$$

- The weights w_{ij} are then associated with the parameter combination $\{\lambda_{n,i}, \beta_j\}$ in the logic tree.

In Equation 4.64, the $q_{i,j}$ term is either taken as 1.0 (or any other constant), or could represent a kernal used to compute a weighted evaluation of the log-likelihood around this node. For example, the full expression could be:

$$w_{ij} = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} q(\lambda_n, \beta \mid \lambda_{n,i}, \beta_j) \ln \mathcal{L}(\lambda_{n,i}, \beta_j) d\lambda_n d\beta}{\sum_j \sum_i \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} q(\lambda_n, \beta \mid \lambda_{n,i}, \beta_j) \ln \mathcal{L}(\lambda_{n,i}, \beta_j) d\lambda_n d\beta \right)} \tag{4.65}$$

Regardless of the form of q_{ij}, whether a constant or some sort of kernel, **there is no mathematical basis for this approach of defining logic tree weights**. In particular, there is no guarantee that the moments of the joint distribution of λ_n and β will be preserved, or that any hazard-relevant derived metric will be preserved.

The basic argument that is used in defence of this approach is that one needs to account for the correlation between λ_n and β. That is true, but there are far better ways to approach this problem, and this will be discussed more completely later in the report in Section 9.

4.3 Problems with the PMLM Approach

The main arguments that are normally offered for adopting the PMLM approach is that there is strong precedent for its use in the UK, that it accounts for variable completeness periods as done by Weichert (1980), and it extends that framework to allow for the penality on the b -value (Dyck, 1985; Veneziano & Dyck, 1985).

However, as noted in Section 2, the PMLM ignores certain issues, with the most important being the uncertainties in the magnitude estimates. What is somewhat surprising is that while authors often cite Musson (2011) regarding the use of the PMLM in the SHARE project, and for its relevance to the UK, they do not seem to pay attention to the section of the report where Musson (2011) discusses the impact of the uncertainties in the magnitude estimates.

This issue will receive more treatment later in the report, but its worth simply outlining the basic issue here. The absolute count of earthquakes in a given catalogue appears in the equations for the optimal λ_n, Equation 4.27, and for the variance of λ_n, Equation 4.43.

The count also appears in the expression we use to solve for β, Equation 4.33, and for the variance of β, Equation 4.45. And, the count appears in the expression for the correlation coefficient, Equation 4.48. So, the absolute count of events in a given catalogue influences all of the parameters that are required to define the logic tree in the hazard analysis. This is unrefutable. This is also not a problem *if* the count is valid. However, if the count is not valid, then there are knock-on consequences for all elements of the logic tree (its centering, its spread, and the correlations among branches and nodes).

Now, consider that the magnitude of an event is not known exactly, but has some nominal uncertainty σ_m. Typically, this uncertainty is assumed to be symmetric (even if not necessarily Gaussian) so that individual events are reported as $m_{\rm obs} \pm \sigma_m$, where $m_{\rm obs}$ is the observed magnitude and σ_m is the uncertainty in the magnitude estimate. The symmetry, and assuming that $m_{\rm obs}$ estimates are not *systematically* biased (so some are above the true mean, others are below, *etc*), implies that roughly half the events will have their true underlying magnitudes above and below the true magnitude. However, we know that the GR distribution is an exponential distribution and that there are far more events below any given magnitude than above it. So, consider the minimum magnitude m_n used in the analysis. There are far more events with true magnitudes that are below m_n that can appear to have $m_{\rm obs} \geq m_n$ than there are events with true magnitudes that are above m_n that appear to have $m_{\rm obs} < m_n$. This means that the observed count of events in a given catalogue, $N_{\rm obs}$, is likely to be an overestimate of the true count of events in that catalogue, $N_{\rm true}$. The extent of the error depends upon the size of the magnitude uncertaitnies, the b-value, and the degree of completeness of the catalogue for magntiudes below m_n – as we will see later in the report.

This effect is known as Eddington bias, and it is a well-known phenomenon in seismology. Indeed, Tinti & Mulargia (1985) proposed a correction to account for this when working with constant magnitude uncertainties. Rhoades (1996) then extended the approach to account for variable magnitude uncertainties. Musson (2011) discusses both of these works and demonstrates with numerical examples how the correction is important. USNRC (2012) introduced two new methods to account for magnitude uncertainties, the M* and N* methods, which are discussed in more detail in Section 6.5.7.

So, the issue is well known, solutions exist, including those formally recognised by the US NRC (USNRC, 2012), yet the PMLM approach ignores this issue entirely. It seems that part of the confusion can be traced to Musson (2012) who discusses various issues related to how magnitude uncertainties are handled, particularly in the context of magnitude conversions. He shows that the use of (outdated) methods for deriving conversion equations can lead to magnitude uncertainties having positive or negative biases. Practitioners seemed to then

justify ignoring the uncertainties by referencing this work. We will see later in this report that this is an important issue that should be properly addressed.

5 Data and Source Zones

Throughout this report use will be made of two distinctly different types of data:

- The actual UK seismicity catalogue, as provided by the BGS (collaborators on this project)
- Synthetic seismicity catalogues, generated using known GR parameters

The UK catalogue adopted here contained all known events up to 2022. It was filtered to remove non-tectonic events (mining, explosions, induced seismicity) prior to analysis. The catalogue includes instrumental and historical events. The objective of the analysis here was not to undertake completely rigorous re-evalutions of the UK catalogue, but to use a real data source in order to test the performance of alternative approaches and to work with a dataset that contains all the real-world sources of noise and any intrinsic departures from the GR model, spatial variations, temporal clustering, etc. In the analyses that are presented throuhgout the report, careful consideration of magnitude and location uncertainty was not given to the historical events. It is very clear that these events have significant magnitude and location uncertainties, and that they are distinctly different from the modern instrumental events in terms of data quality. The choice to not pay close attention to these events does not reflect the view of the author that these are unimportant. Rather, it is clear that these events are very important and the large uncertainties associated with them must be considered carefully in real analyses. Instead, the choice is just for practical convenience because it is possible to demonstrate the core issues with existing practice without taking special care of the historical events. It is important to note that the uncertainties in GR parameters computed on the basis of the treatment of the historical events in this study are lower bounds on the true uncertainties.

While it is useful to work with real data and to compare differences in appraoches when using this data, and the limited samples sizes that one obtains when partitioning the data over different source zones, the core problem is that we do not know what the ‘true’ answers are for each type of analysis. So, working with real data allows one to identify differences in the results obtained from different methods, but it does not allow for any assessment of bias.

For that reason, a lot of work is done using synthetic catalogues. The generation of these catalogues is designed to mimic the generative process relevant for the real seismicity, but starts from a known GR model. We can then generate multiple synthetic catalogues, each with different realizations of the GR model, and compare the results obtained from different methods. This allows us to assess the bias in the results obtained from different methods, and to compare the performance of different methods in a more rigorous way.

5.1 UK Earthquake Catalogue

The analyses in this report use the BGS earthquake catalogue provided for this project by Ilaria Mosca (BGS), containing all known events within the geographic bounds of latitude 49.0°–61.0° N and longitude 8.0° W–3.0° E, up to 31 December 2022. The full catalogue contains 15,337 events spanning from 1122 to 2022.

5.1.1 Catalogue fields and event classification

The catalogue provides the following fields for each event: date and time of occurrence, location (latitude, longitude, depth), local magnitude (M_L) with associated standard deviation, magnitude type, distance type, event type classification, quality factor, and a location description. Table 5.1 summarises the breakdown of events by type classification.

Table 5.1: Breakdown of events in the BGS catalogue by event type classification.

Type	Count	%	Description
Tectonic	7,902	51.5	No type flag (blank)
Mining/collapse	3,627	23.6	C flag
Explosion/quarry blast	3,111	20.3	E flag
Induced	695	4.5	I flag

Nearly half the raw catalogue consists of non-tectonic events; these are filtered out prior to all analyses presented here, leaving 7,902 tectonic events.

5.1.2 Instrumental versus historical events

The distinction between instrumental and historical events is important because it directly affects the quality of both location and magnitude estimates. Events prior to the establishment of the modern seismic network (roughly 1970) rely on macroseismic intensity reports from which magnitudes are estimated. Of the 7,902 tectonic events, the vast majority (approximately 93%) date from 1970 onwards and are instrumentally recorded. Only 574 tectonic events (7.3%) pre-date 1970, and these are concentrated at larger magnitudes where felt reports provide the primary evidence. Table 5.2 shows the magnitude distribution for tectonic events only.

Table 5.2: Distribution of tectonic events by magnitude range.

Magnitude range	Count	% of tectonic
$M_L < 2.0$	6,234	78.9
$2.0 \leq M_L < 3.0$	1,173	14.8
$3.0 \leq M_L < 4.0$	354	4.5
$4.0 \leq M_L < 5.0$	108	1.4
$5.0 \leq M_L < 6.0$	30	0.4
$M_L \geq 6.0$	3	< 0.1

The table makes clear that the catalogue is overwhelmingly composed of small events. For typical analysis thresholds of $M_L \geq 3.0$, we have fewer than 500 tectonic events across the entire UK; at $M_L \geq 4.0$, roughly 140; and at $M_L \geq 5.0$, only 33. These numbers highlight the fundamental challenge for seismic hazard analysis: the events that contribute most to hazard at critical infrastructure sites are precisely those for which we have the fewest observations.

5.1.3 Quality factor for historical events

The BGS catalogue includes a quality factor for historical events based on macroseismic data. This factor uses a two-character code where the capital letter indicates the maximum error in epicentre location and the lowercase letter indicates the quality of the underlying data/reports:

Table 5.3: Quality factor classification for historical events in the BGS catalogue.

Code	Epicentre error	Code	Data quality
A	< 5 km	a	Good
B	5–15 km	b	Fair
C	15–30 km	c	Poor
D	> 30 km	d	Poor

Of the 7,902 tectonic events, 472 (6.0%) carry a quality factor. Among the 495 tectonic events with $M_L \geq 3.0$—the magnitudes most relevant for hazard analysis—262 (52.9%) have quality factors assigned. The remaining events are either modern instrumental recordings (where formal uncertainties are provided directly) or historical events for which the quality factor was not included in the files provided. Table 5.4 shows the distribution of quality factor codes among tectonic events that have them.

Table 5.4: Distribution of quality factors among the 477 events with quality codes.

Quality factor	Count	Interpretation
Aa	128	Well-located, good data
Ba	202	Moderate location, good data
Ab	30	Well-located, fair data
Bb	51	Moderate location, fair data
Ca	29	Poor location, good data
Cb	13	Poor location, fair data
Da	12	Very poor location, good data
Db	4	Very poor location, fair data
Other	8	Including Cc, Dc, Bd, and anomalous entries

The majority of quality-coded events fall in the Aa and Ba categories, indicating that the historical events retained in the catalogue are generally those with the best available macroseismic evidence. Nevertheless, even "good quality" historical events have location uncertainties of up to 15 km and magnitude uncertainties that are substantially larger than for instrumental events.

5.1.4 Magnitude uncertainties

The BGS catalogue provides standard deviations of M_L for instrumentally-recorded events. Events with unknown or unavailable uncertainties are coded as 999.9; this applies to all historical events and all events after 2022 (where uncertainties had not yet been computed at the time the catalogue was provided). Of the 15,337 events in the full catalogue, 5,187 (33.8%) have reported M_L uncertainties. Among these, the median uncertainty is 0.2 and the mean is 0.26, with an interquartile range of 0.1 to 0.4.

Figure 5.1 shows the distribution of reported M_L uncertainties for instrumental events, grouped into 5-year windows from 1970 to 2020. The distributions are well-described by Gamma distributions, and show a clear temporal evolution: uncertainties generally decreased from the mid-1970s (mean $\approx$ 0.30) through the late 1990s (mean $\approx$ 0.19), before increasing again from 2000 onwards (mean $\approx$ 0.29 in 2005–2019). This later increase likely reflects the inclusion of smaller events with larger relative uncertainties as detection capability improved, rather than any degradation of the seismic network itself.

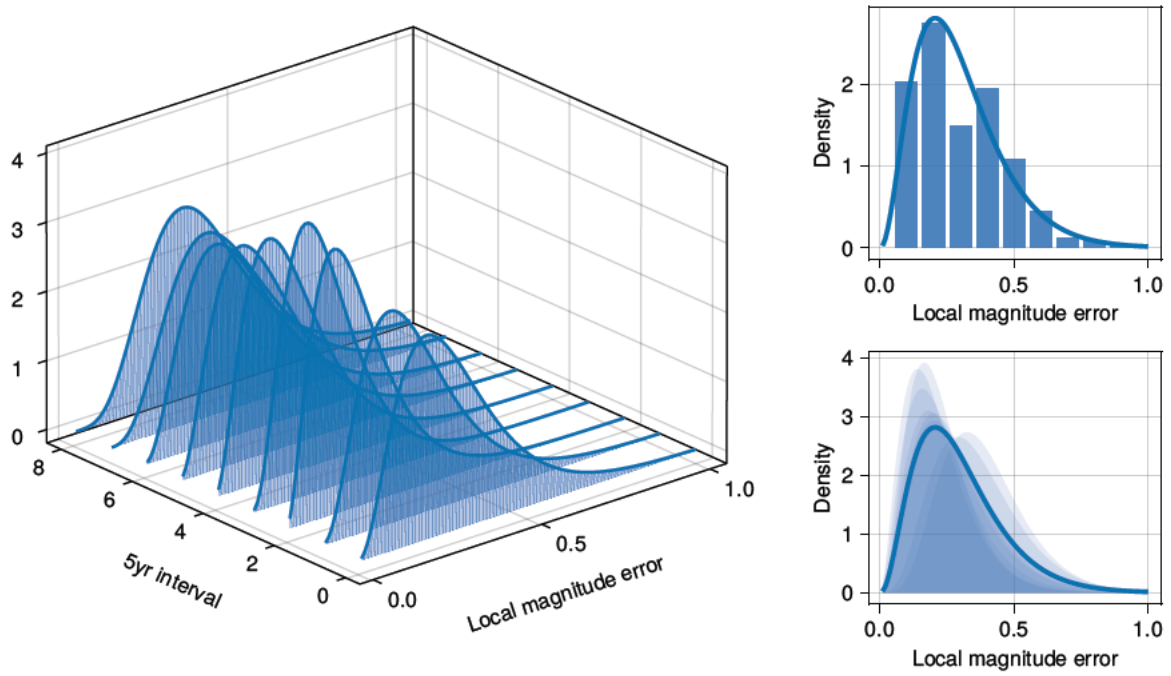

Figure 5.1: Distribution of reported M_L uncertainties in the BGS catalogue. Left: Gamma distributions fitted to 5-year windows from 1970 to 2020, showing temporal evolution. Top right: overall histogram with fitted Gamma distribution. Bottom right: overlay of all 5-year window distributions showing the range of variation.

For events without reported uncertainties (primarily historical events), default values are assigned based on era, following standard practice:

Table 5.5: Default M_L measurement uncertainties assigned by era when catalogue values are unavailable.

Period	σ_{M_L}	Basis
Pre-1900	0.5	Historical macroseismic estimates
1900–1970	0.4	Early instrumental
1970–1990	0.25	Improved seismic network
Post-1990	0.15	Modern network

These measurement uncertainties represent only one component of the total magnitude uncertainty relevant for seismic hazard analysis. As discussed in detail in Section 6, the total uncertainty in moment magnitude M_w combines three independent sources: the measurement uncertainty in M_L (propagated through the conversion relationship), the aleatory variability of the M_L-to-M_w conversion itself, and rounding uncertainty. For the Grünthal et al. (2009) conversion adopted here, the conversion uncertainty alone is $\sigma \approx 0.26$–0.39 depending on magnitude (see Table 7.1), meaning that for modern instrumental events with

$\sigma_{M_L} \approx 0.15$, the conversion uncertainty is the dominant contributor to total M_w uncertainty. For historical events with $\sigma_{M_L} \approx 0.5$, the measurement uncertainty becomes comparable to or exceeds the conversion uncertainty, resulting in total M_w uncertainties of approximately 0.55–0.65.

Figure 5.2 shows how the cumulative number of earthquakes in the UK catalogue changes over time, for various minimum magnitude thresholds. For typical analysis thresholds of $M_L \geq 2.5$ we can see that the entire catalogue has fewer than 1000 events. This is a very limited sample size for statistical analyses (further supporting the use of synthetic catalogues), and it is also important to keep in mind that our end objective is to define rates of earthquake occurrence for events that contribute to seismic hazard for critical infrastructure. If we have fewer than 1000 events for $M_L \geq 2.5$, and a typical b-value around 1.0, then we have fewer than 100 events for $M_L \geq 3.5$, and fewer than 10 events for $M_L \geq 4.5$. These numbers are distributed across the entire UK, so the expectation is that we have very few, if any, events at magnitudes of relevance for hazard in many source zones.

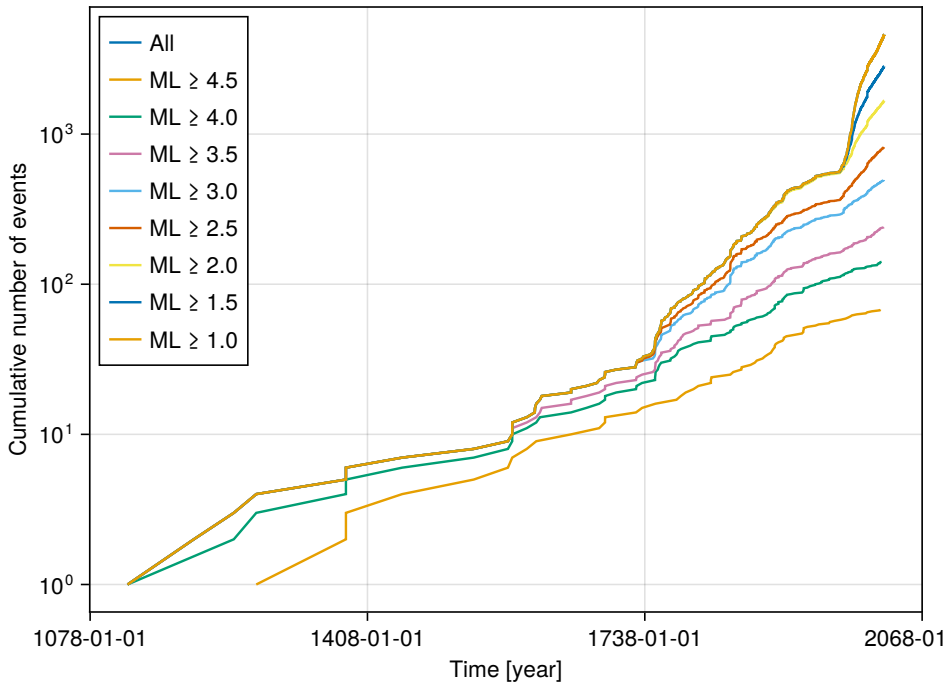

Figure 5.2: Cumulative number of earthquakes in the UK catalogue over time, shown for various minimum magnitude thresholds.

5.2 Catalogue Completeness

Completeness levels follow Musson (2007) with magnitude thresholds and corresponding complete periods:

Table 5.6: Completeness periods for different magnitude thresholds and regions.

M_L	UK	SE England	Dogger	Viking
3.0	1970	1970	1970	1970
3.5	1850	1850	1970	1970
4.0	1750	1750	1850	1970
4.5	1700	1700	1750	1900
5.0	1650	1650	1650	1900
5.5	1650	1300	1650	1900
6.5	1000	1000	1000	1700

In other analyses, the completeness limits are extended to smaller magnitudes, reflecting the more realistic levels of completeness identified by Mosca et al. (2022).

5.3 Declustering

Dependent events (foreshocks and aftershocks) are identified and removed using the window method of Burkhard & Grünthal (2009) with manual adjustments for known sequences (Torridon 1934-36, Kintail 1974-75, Lleyn Peninsula 1984-95). This approach was not adopted as any sort of endorsement of the method itself, but simply because it represented what has been done in previous analyses of UK seismicity. Declustering is a complex topic, and there are many different methods that can be used to identify and remove dependent events. But, the UK catalogue is too small to allow for any sort of rigorous assessment of the impact of declustering on the results of analyses. The focus in the present report is more about identifying optimal GR parameter calibration approaches *given* a particular catalogue that is assumed to be valid and realistic.

5.4 Source Zone Definitions

The 22 UK source zones defined in Mosca et al. (2022) are used, with areas calculated via transformation to British National Grid coordinates. The basic geometry of the source zones are shown in Figure 5.3, along with the seismicity. The source zone vertex coordinates are available from the BGS website and are defined in Mosca et al. (2022).

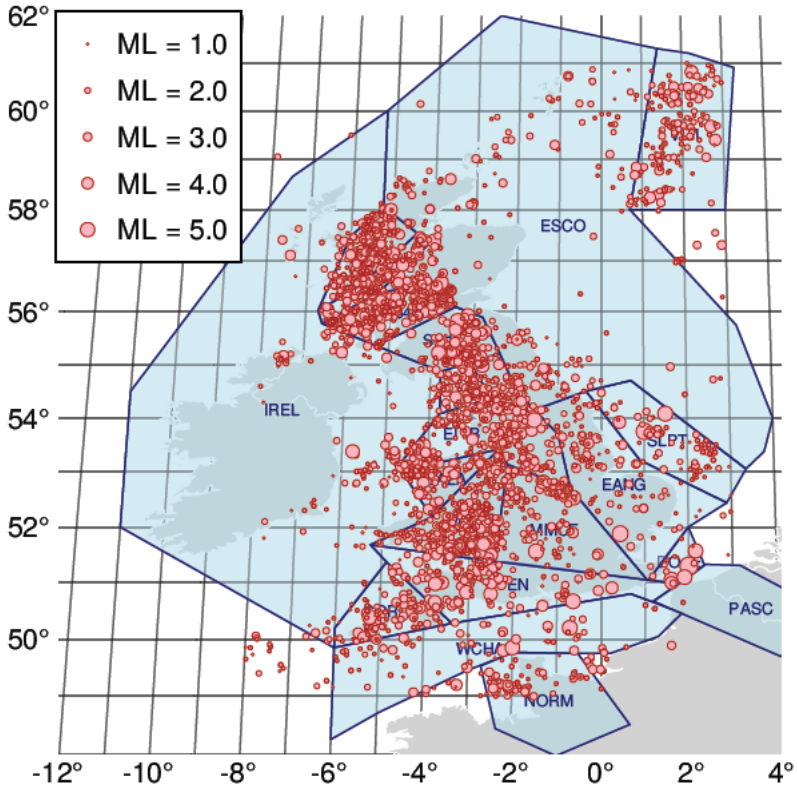

Figure 5.3: UK source zones with declustered earthquake locations. Marker size scaled by magnitude.

Like with the declustering approach, and the specfication of the completeness levels, the use of these nominal values does not constitute an endorsement of their validity. They are simply adopted on the basis of precedent to enable us to work with a realistic catalogue that possesses the similar characteristics to what practitioners typically work with. As will be seen later in the report, one can very easily make strong statistical arguments that would suggest modifications of the source zone boundaries, or indeed that the source zones should be defined in a completely different way.

5.5 Synthetic Catalogue Generation

While the UK catalogue provides a realistic test bed, it does not allow us to assess the bias of any estimation method because the true GR parameters are unknown. For that purpose, synthetic catalogues are generated with known parameters, allowing rigorous assessment of both bias and uncertainty calibration across methods. The generation procedure is designed to faithfully replicate the physical processes that produce an observed earthquake catalogue, including all the sources of noise that affect real data.

5.5.1 Generative model

The synthetic catalogue generation follows a forward simulation that mirrors the assumed physical process from true earthquake occurrence through to reported magnitude:

1. **True event occurrence:** The number of events above a generation floor magnitude m_{floor} over an observation period T is drawn from a Poisson distribution: $N \sim \text{Poisson}(\lambda_{\text{floor}} \cdot T)$, where $\lambda_{\text{floor}} = \lambda_{m_{\min}} \cdot 10^{b(m_{\min} - m_{\text{floor}})}$ extrapolates the target rate down to m_{floor}.
2. **True magnitudes:** Each event's true moment magnitude M_w^{true} is sampled from the truncated exponential (Gutenberg-Richter) distribution on $[m_{\text{floor}}, m_{\max}]$:

$$M_w^{\text{true}} = m_{\text{floor}} - \frac{1}{\beta} \ln\left(1 - U \cdot \left(1 - e^{-\beta(m_{\max} - m_{\text{floor}})}\right)\right) \tag{5.1}$$

   where $U \sim \text{Uniform}(0, 1)$ and $\beta = b \ln 10$.
3. **Magnitude conversion:** True M_w values are converted to the local magnitude scale using the inverse of the Grünthal et al. (2009) quadratic relationship to obtain M_L^{true}.
4. **Conversion uncertainty:** The inherent scatter in the M_w–M_L relationship is applied: $M_L^{\text{conv}} = M_L^{\text{true}} + \varepsilon_{\text{conv}}$, where $\varepsilon_{\text{conv}} \sim \mathcal{N}(0, \sigma_{\text{conv}}(M_L))$ and σ_{conv} is the magnitude-dependent Grünthal conversion uncertainty transformed to the M_L scale (see Section 6 for details).
5. **Measurement error:** Instrumental measurement noise is added: $M_L^{\text{meas}} = M_L^{\text{conv}} + \varepsilon_{\text{ML}}$, where $\varepsilon_{\text{ML}} \sim \mathcal{N}\left(0, \sigma_{M_L}\right)$.
6. **Rounding:** The measured magnitude is rounded to the reporting precision (typically $\Delta m = 0.1$): $M_L^{\text{reported}} = \text{round}(M_L^{\text{meas}} / \Delta m) \times \Delta m$.
7. **Catalogue selection:** Events enter the catalogue if M_L^{reported} falls at or above the effective detection threshold, which is computed from the target minimum M_w analysis magnitude accounting for the conversion and rounding.
8. **Analysis window:** Events are retained for analysis if their converted M_w falls within $[m_{\min}, m_{\max}]$.

The generation floor m_{floor} is set well below the analysis threshold $m_{\min}$ (typically $m_{\text{floor}} = 1.0$ versus $m_{\min} = 3.0$). This ensures that the catalogue includes events whose true magnitudes are below $m_{\min}$ but whose observed magnitudes scatter above it (scatter-IN), and conversely that events truly above $m_{\min}$ that scatter below the threshold are properly lost (scatter-OUT). This mechanism is central to the Eddington bias phenomenon that the Full Bayesian model is designed to handle.

5.5.2 Default parameters

Unless stated otherwise, synthetic catalogues are generated with the following default parameters:

Table 5.7: Default parameters for synthetic catalogue generation.

Parameter	Default	Description
b	1.0	Gutenberg-Richter slope
$\lambda(\geq m_{\min})$	2.0 yr^{-1}	Rate above analysis threshold
T	50 yr	Observation period
$m_{\min}$	3.0	Minimum analysis magnitude (M_w)
$m_{\max}$	6.5	Maximum magnitude (M_w)
m_{floor}	1.0	Generation floor (M_w)
σ_{M_L}	0.25	Measurement uncertainty
Δm	0.1	Reporting precision
σ_{conv}	Grünthal	Magnitude-dependent conversion uncertainty

These values are representative of typical UK source zones. For example, $\lambda = 2.0$ yr^{-1} above $M_w = 3.0$ with $T = 50$ yr yields roughly 100 observed events—similar to the larger UK source zones. The measurement uncertainty $\sigma_{M_L} = 0.25$ is close to the median of reported uncertainties in the BGS catalogue for the instrumental period (see Section 5.1.4).

5.5.3 Monte Carlo validation protocol

To assess method performance, the generation procedure is repeated n_{rep} times (typically 20–100 replications) with different random seeds but identical true parameters. Each synthetic catalogue is then analysed using all methods under comparison (Weichert MLE, PMLM, the Full Bayesian model, etc.). This allows computation of:

- **Bias:** the systematic difference between the mean estimate and the true value across replications
- **Posterior SD calibration:** the ratio of between-replication standard deviation to mean within-replication posterior standard deviation (a ratio near 1.0 indicates well-calibrated uncertainties)
- **Coverage:** the fraction of replications for which the true value falls within the stated credible interval

This protocol is used extensively in Section 6 to validate the Full Bayesian model and to quantify the biases inherent in classical methods.

6 Theme A: Full Bayesian Bayesian Model

This section presents the comprehensive Bayesian framework for estimating magnitude-frequency parameters – termed the "Full Bayesian" model. The model addresses the fundamental challenges that arise when magnitudes are uncertain, selection is threshold-based, and completeness varies with magnitude.

6.1 The Physical Setting

6.1.1 What We Observe vs. What We Want

Earthquake catalogues record **observed** magnitudes m^{obs}, but seismic hazard analysis requires the distribution of **true** magnitudes m^{true}. The relationship between these involves multiple sources of uncertainty:

Table 6.1: Sources of magnitude uncertainty in UK catalogues

Source	Typical Value	Effect
Measurement error (σ_{ML})	≈ 0.2–0.3	Random scatter in reported ML
Conversion uncertainty (σ_{conv})	≈ 0.25–0.30	Scatter in ML→Mw relationship
Rounding (σ_{round})	≈ 0.029	Discretisation to 0.1 units

The catalogue selection process introduces Eddington bias. As noted earlier in the report, because the Gutenberg-Richter distribution is exponentially decreasing, more events scatter into the catalogue from below the threshold than scatter out from above it. This means that event counts of *reported* magnitudes are systematically biased towards higher values. This has been recognised for a long time, with Tinti & Mulargia (1985) and Dyck (1985) both addressing this over 40 years ago.

6.1.2 Variable Completeness

Historical earthquake catalogues have variable completeness: larger events are recorded over longer time periods because they were felt over wider areas and documented in historical records. This means that an observed large earthquake ($M_w \approx 5$) could have occurred and been recorded at any time in the past ~300 years, while a small earthquake ($M_w \approx 3$) is far more likely to have been recorded within the past ~50 years, unless it occurred very close to a population centre. The observation period $T(m)$ is a function of true magnitude.

The actual completeness is a function of the network characteristics as well as the event size. For historical events, the 'network' is really the spatial distribution of population

centres. For instrumental catalogues, the network is the seismic network, and the completeness is a function of the distance from the epicentre to the nearest station as well as the sensitivity of the instruments at each station and the background noise. Mosca et al. (2022) refer to work (the report accompanying their paper) that shows the spatial variation of complteness levels (or detection thresholds) for the UK instrumental catalogue as a function of time. That analysis shows that magnitudes of completeness that are typically assumed are probably conservative.

6.2 The Generative Model

The Full Bayesian model is fully Bayesian and its often helpful to think of it as a data generating process. This process mimics the way that the synthetic catalogues are generated and allows one to understand where uncertainties enter the process and which latent variables, prior assumptions, and modelling choices are most important.

The Full Bayesian model is defined by the following data-generating process:

Step 1: For each completeness bin j with observation period T_j:

Generate event count: $N_j \sim \text{Poisson}(\lambda_{\text{floor}} \cdot T_j \cdot p_j)$

where $p_j = \int_{\text{bin}_j} f_{\text{GR}}(m \mid \beta)\, \mathrm{d}m$ is the GR probability for bin j

Step 2: For each event i:

(a) Draw true M_w from truncated GR within its bin:

$M_w^{\text{true}} \sim f_{\text{GR}}(\cdot \mid \beta, \text{bin}_j)$

(b) Convert to local magnitude with aleatory scatter:

$M_L^{\text{true}} = g^{-1}(M_w^{\text{true}}) + \varepsilon_{\text{conv}}, \quad \varepsilon_{\text{conv}} \sim \mathcal{N}(0, \sigma^2_{\text{conv}})$

(c) Add measurement error:

$M_L^{\text{meas}} = M_L^{\text{true}} + \varepsilon_{\text{ML}}, \quad \varepsilon_{\text{ML}} \sim \mathcal{N}(0, \sigma^2_{\text{ML}})$

(d) Apply rounding:

$M_L^{\text{rep}} = \text{round}(M_L^{\text{meas}}, \Delta m)$

Step 3: Selection

Event enters catalogue if $M_L^{\text{rep}} \geq m_L^{\text{threshold}}$

6.3 The Thinned Poisson Likelihood

An earthquake catalogue is a thinned Poisson process in observed magnitude space. The original (latent) process generates events according to the GR distribution in true magnitude space. The thinning arises from the probabilistic mapping from true to observed magnitude and the threshold-based selection.

6.3.1 The Per-Event Intensity

In a standard Poisson process, each observed event contributes to the likelihood through a quantity called its *intensity*. In plain terms, the intensity of an event is a measure of how likely it was that the model would produce an observation at that particular value. A high intensity means the model readily explains this observation; a low intensity means the observation is surprising. In the familiar case with no magnitude uncertainty (i.e., the Weichert or PMLM framework), the intensity of an event observed at magnitude m_i is simply the GR density at that magnitude multiplied by how long we have been watching for events of that size: $I_i = T(m_i) \cdot f_{\mathrm{GR}}(m_i \mid \beta)$. This is directly analogous to the expected count in a magnitude bin – but evaluated at a single observed value rather than integrated over a bin.

When magnitudes are uncertain, we do not know the true magnitude that produced the observation m_i. A range of true magnitudes could have given rise to this observation, each with a different probability. The intensity must therefore sum up (integrate over) the contributions from all possible true magnitudes, weighting each by how likely the GR model says it is, how long we have been watching for events at that true magnitude, and how likely the observation m_i is given that particular true value:

$$I_i = \int_{m_{\mathrm{floor}}}^{m_{\max}} T(m_w) \cdot f_{\mathrm{GR}}(m_w \mid \beta) \cdot p(m_i \mid m_w)\, \mathrm{d}m_w \tag{6.1}$$

where:

- $T(m_w)$ is the observation period for true magnitude m_w
- $f_{\mathrm{GR}}(m_w \mid \beta)$ is the GR density at true magnitude m_w
- $p(m_i \mid m_w)$ is the probability of observing m_i given true magnitude m_w

When $\sigma \to 0$, the observation model $p(m_i \mid m_w)$ becomes a spike at $m_w = m_i$ and the integral collapses back to $T(m_i) \cdot f_{\mathrm{GR}}(m_i \mid \beta)$ — the familiar expression with no uncertainty. The integral is thus a direct generalisation that reduces to the standard case.

The presence of $T(m_w)$ inside the integral requires careful justification. Consider an event observed at $m_L^{\mathrm{obs}} = 4.5$. This could have arisen from a true $M_w = 4.0$ event (observable for $T \approx 270$ years) that scattered upward, or a true $M_w = 5.0$ event (observable for $T \approx 370$ years) that scattered downward. The relative likelihood of these two explanations depends not only on the GR distribution (which favours the smaller magnitude), but also on how long each scenario had to occur. The $T(m_w)$ factor captures this exposure-time weighting.

6.3.2 The Expected Count

The expected total count integrates the intensity over all observed magnitudes, accounting for the GR distribution, selection probability, and observation period:

$$\Lambda = \lambda_{\text{floor}} \cdot \sum_{j=1}^{J} T_j \cdot \int_{\text{bin}_j} f_{\text{GR}}(m \mid \beta) \cdot P(\text{selected} \mid m)\, \mathrm{d}m \tag{6.2}$$

where the selection probability is:

$$P(\text{selected} \mid m) = \Phi\left(\frac{\mu_L(m) - m_L^{\text{threshold}}}{\sigma_{\text{total}}}\right) \tag{6.3}$$

The integration must include the scatter-IN region: magnitudes below the nominal threshold ($m < m_{\text{min}}$) that can still enter the catalogue by scattering above the threshold. For this region, we use $T = T_1$ (the observation period of the lowest completeness bin).

6.3.3 The Complete Log-Likelihood

$$\ln \mathcal{L}(\beta, \lambda_{\text{floor}}) = \underbrace{\sum_{i=1}^{N} \ln I_i}_{\substack{\text{Per-event} \\ \text{intensity}}} + \underbrace{N \cdot \ln \lambda_{\text{floor}}}_{\substack{\text{Poisson} \\ \text{count}}} - \underbrace{\Lambda}_{\substack{\text{Expected} \\ \text{count}}} \tag{6.4}$$

Important: Consistency Requirement: The $T(m)$ weighting must appear in **both** I_i and Λ. This follows directly from the Poisson process likelihood. The expected count tells us how many events we expect to observe; the per-event intensities tell us the relative probability of observing each event. Both must use the same weighting.

6.4 Why Marginalisation is Essential

6.4.1 The Latent Variable Approach

In a fully Bayesian framework, a common approach is to explicitly model the true underlying magnitude values as parameteres to be predicted – like in a missing data problem. This intuitive approach treats the true magnitudes $\{m_i^{\text{true}}\}$ as latent parameters to be estimated alongside the GR parameters. This formulation has $N + 2$ parameters: the N true magnitudes plus β and λ_{floor}. While conceptually appealing, this approach has serious practical problems.

6.4.2 The Identifiability Problem

The latent variable approach suffers from severe identifiability issues. Different combinations of $(\beta, \{m_i^{\text{true}}\})$ produce nearly identical likelihoods: if β increases, the model can compensate by shifting all m_i^{true} values upward. This creates a ridge-like posterior geometry that is extremely difficult for MCMC samplers, leading to convergence failures, poor $\hat{R}$ diagnostics, and overconfident uncertainty estimates.

6.4.3 The Marginalisation Approach

To overcome the issues of identifiability, we can instead integrate out the latent magnitudes analytically. This approach is known as marginalisation and is exact; it is not some approximation. By integrating out the latent magnitudes analytically, we obtain a likelihood that depends only on $(\beta, \lambda_{\text{floor}})$:

$$p\left(m_i^{\text{obs}} \mid \beta, \lambda_{\text{floor}}\right) = \lambda_{\text{floor}} \cdot \int f_{\text{GR}}(m \mid \beta) \cdot p\left(m_i^{\text{obs}} \mid m\right) \mathrm{d}m = \lambda_{\text{floor}} \cdot I_i \tag{6.5}$$

This marginalisation eliminates N nuisance parameters, leaving only the 2 parameters of scientific interest.

Equivalence Theorem: The marginal posterior for (β, λ) is identical whether we sample latent magnitudes directly or integrate them out analytically. Marginalisation is not an approximation – it is exact. The only difference is computational: we perform the integration analytically (via numerical quadrature) rather than through MCMC sampling, achieving dramatically better computational properties.

6.5 Recovery of Classical Methods as Limiting Cases

A critical validation of the Full Bayesian framework is demonstrating that well-established classical methods emerge as special cases when appropriate simplifying assumptions are made. The framework just presented is the most general form of the model, and its not immediately apparent how it relates to previous methods. The present section provides the translation which is essential so that practitioners can trust that the Full Bayesian captures everything previous approaches attempted. The Bayesian framework is not something foreign to existing practice — it is a strict generalisation that reduces exactly to familiar methods under the assumptions those methods make. The key message is that every classical method listed below is already embedded within the Full Bayesian; the only question is how many of the simplifying assumptions one is willing to impose.

To make this concrete, recall the three key equations of the Full Bayesian from the previous section. The per-event intensity (Equation 6.1):

$$I_i = \int_{m_{\text{floor}}}^{m_{\max}} T(m_w) \cdot f_{\text{GR}}(m_w \mid \beta) \cdot p(m_i \mid m_w) \, \mathrm{d}m_w \tag{6.6}$$

the expected count (Equation 6.2):

$$\Lambda = \lambda_{\text{floor}} \cdot \sum_{j=1}^{J} T_j \cdot \int_{\text{bin}_j} f_{\text{GR}}(m \mid \beta) \cdot P(\text{selected} \mid m) \, \mathrm{d}m \tag{6.7}$$

and the complete log-likelihood (Equation 6.4):

$$\ln \mathcal{L} = \sum_{i=1}^{N} \ln I_i + N \ln \lambda_{\text{floor}} - \Lambda \tag{6.8}$$

Every classical method can be obtained by progressively stripping away components of these equations.

6.5.1 Aki (1965): The Simplest Case

The Aki (1965) MLE is the simplest possible estimator for the GR b-value. It assumes: (i) no magnitude uncertainty ($\sigma = 0$); (ii) a single, constant observation period T for all magnitudes; and (iii) it does not estimate the rate — only β.

Under these three assumptions, the Full Bayesian simplifies as follows. Setting $\sigma = 0$ collapses the observation model to a Dirac delta, $p(m_i \mid m_w) \to \delta(m_i - m_w)$, so the per-event intensity reduces to:

$$I_i \to T \cdot f_{\text{GR}}(m_i \mid \beta) \tag{6.9}$$

Since T is constant, it factors out of the log-likelihood. The expected count becomes $\Lambda = \lambda_{\text{floor}} \cdot T$, and the $N \ln \lambda_{\text{floor}}$ term together with Λ yields the standard Poisson count component. Maximising with respect to β alone (treating N as given), the log-likelihood reduces to:

$$\ln \mathcal{L} \propto \sum_{i=1}^{N} \ln f_{\text{GR}}(m_i \mid \beta) \tag{6.10}$$

This is precisely the multinomial likelihood used by Aki (1965). Setting its derivative to zero gives the well-known MLE:

$$\hat{\beta}_{\text{Aki}} = \frac{1}{\overline{m} - m_n} \quad \text{where} \quad \overline{m} = \frac{1}{N} \sum_{i=1}^{N} m_i \tag{6.11}$$

Note that the Aki estimator treats the observed event count N as fixed and does not attempt to estimate the activity rate λ_n. The Full Bayesian, in contrast, jointly estimates both λ_{floor} and β — the rate information is lost if one conditions on N.

6.5.2 Weichert (1980): The $\sigma \to 0$ Limit

The Weichert (1980) method generalises Aki by allowing variable completeness periods across magnitude bins while still assuming no magnitude uncertainty. This is the foundation upon which the PMLM is built.

Setting $\sigma = 0$ in the Full Bayesian, the observation model again collapses to a Dirac delta function:

$$\lim_{\sigma \to 0} p(m_i \mid m_w) = \delta(m_i - m_w) \tag{6.12}$$

and the selection probability becomes a step function: $P(\text{selected} \mid m) \to \mathbf{1}(m \geq m_n)$.

Per-event intensity. The integral in Equation 6.1 collapses via the sifting property of the delta function:

$$I_i \to T(m_i) \cdot f_{\mathrm{GR}}(m_i \mid \beta) \tag{6.13}$$

The key point is that $T(m_i)$ is now the observation period at the *observed* magnitude — there is no integration over possible true magnitudes because there is no uncertainty.

Expected count. With $P(\text{selected} \mid m) = 1$ for $m \geq m_n$ and the magnitude range partitioned into completeness bins, Equation 6.2 becomes:

$$\Lambda \to \lambda_n \cdot \sum_{j=1}^{J} T_j \cdot \int_{\mathrm{bin}_j} f_{\mathrm{GR}}(m \mid \beta)\, \mathrm{d}m = \lambda_n \sum_{j=1}^{J} T_j \gamma_j(\beta) \tag{6.14}$$

where $\gamma_j(\beta)$ is the GR bin probability defined in Equation 3.32.

Log-likelihood. Substituting into Equation 6.4:

$$\ln \mathcal{L} = \sum_{i=1}^{N} \left[\ln T(m_i) + \ln f_{\mathrm{GR}(m_i \mid \beta)}\right] + N \ln \lambda_n - \lambda_n \sum_{j=1}^{J} T_j \gamma_j(\beta) \tag{6.15}$$

But we can equivalently work with the binned form. If events are grouped into magnitude bins with n_j events in bin j, each experiencing observation period T_j, the log-likelihood becomes:

$$\ln \mathcal{L} = \sum_{j=1}^{J} \left[n_j \ln\left(\lambda_n \gamma_j T_j\right) - \lambda_n \gamma_j T_j\right] + \mathrm{const} \tag{6.16}$$

This is *exactly* the binned Poisson likelihood of Equation 3.35 from Section 3.4, and it is exactly the Weichert likelihood. The correspondence is term-by-term:

Table 6.2: Term-by-term correspondence between the Full Bayesian (with $\sigma \to 0$) and the Weichert/ PMLM binned Poisson likelihood.

Full Bayesian ($\sigma \to 0$)	Weichert / PMLM notation	Physical meaning
$I_i = T(m_i) \cdot f_{\mathrm{GR}(m_i)}$	$t_i \cdot \gamma_i(\beta)$	Exposure-weighted density per event
$\Lambda = \lambda_n \sum_j T_j \gamma_j$	$\lambda_n \sum_i \gamma_i t_i$	Expected total count
$\sum_i \ln I_i + N \ln \lambda_n$	$n_T \ln \lambda_n + \sum_i n_i \ln(\gamma_i t_i)$	Event contribution to log-likelihood
λ_{floor}	λ_n	Rate at minimum magnitude

The crucial point is that the Full Bayesian and the Weichert method *use the same likelihood* when magnitude uncertainty is absent. They are not different models — they are the same model. The only difference is that the Full Bayesian additionally allows for magnitude uncertainty through the observation model $p(m_i \mid m_w)$ and the integration in Equation 6.1.

6.5.3 Johnston PMLM: Adding the Prior

The Penalised Maximum Likelihood Method (Johnston et al., 1994) extends Weichert by adding a Gaussian penalty on β to stabilise estimates for small samples. As shown in Section 4, the PMLM log-likelihood is (Equation 3.35 with penalty):

$$\ln \mathcal{L}_{\text{PMLM}} = \underbrace{\sum_{i=1}^{I} [n_i \ln(\lambda_n \gamma_i t_i) - \lambda_n \gamma_i t_i]}_{\text{Weichert likelihood}} - \underbrace{\frac{W}{2}(\beta - \beta_p)^2}_{\text{Prior penalty}} \tag{6.17}$$

In the Bayesian framework, this penalty term is *exactly* a Gaussian prior $\beta \sim \mathcal{N}(\beta_p, 1/W)$. Therefore:

$$\{\text{Full Bayesian}\} + \{\sigma = 0\} + \{\mathcal{N}(\beta_p, 1/W) \text{ prior}\} = \{\text{PMLM}\} \tag{6.18}$$

The PMLM is not merely *similar* to a Bayesian method — it *is* a Bayesian method (specifically, MAP estimation). As discussed in Section 4, the conventional UK weight $W = 25$ corresponds to $\sigma_\beta = 0.2$ or equivalently $\sigma_b \approx 0.087$, meaning that the prior constrains b to lie within approximately ± 0.17 of the prior mean (at the 95% level). For UK source zones with typical sample sizes of 30–100 events, Table 6.3 shows the effective influence of this prior on the final estimate.

Table 6.3: Effective weight of the PMLM prior ($W = 25$) relative to data as a function of sample size.

N events	Data weight	Prior weight	$\hat{b}_{\text{PMLM}}$ vs $\hat{b}_{\text{MLE}}$
25	50%	50%	Midpoint of MLE and prior
50	67%	33%	Closer to MLE
100	80%	20%	Dominated by data
200	89%	11%	Prior nearly irrelevant

For UK source zones with typical $N \approx 30$–100 events, the prior has significant influence on the b-value estimate. This is by design (the prior prevents extreme b-value estimates in data-poor zones), but it is important for practitioners to understand that the PMLM is making a specific prior assumption, and that this assumption can be made explicit and modified within the Bayesian framework.

The Full Bayesian generalises the PMLM in three ways: (i) it incorporates magnitude uncertainty rather than assuming $\sigma = 0$; (ii) it returns the full posterior distribution rather than just the MAP point estimate; and (iii) it allows arbitrary priors rather than being restricted to a Gaussian on β alone.

6.5.4 Tinti & Mulargia (1985): Rate Bias

Tinti & Mulargia showed that magnitude uncertainty causes systematic overestimation of activity rates. The ratio of observed to true rate above any threshold is:

$$\frac{N(\text{observed})}{N(\text{true})} = \exp\left(\beta^2 \frac{\sigma^2}{2}\right) \tag{6.19}$$

For typical UK conditions ($b = 1.0$, $\sigma = 0.30$): $\exp(\beta^2\sigma^2/2) = 1.27$ (27% rate overestimate).

This bias emerges naturally from our expected count integral. The Tinti-Mulargia correction is automatically incorporated in the Full Bayesian model.

Important: Critical Limitation of Tinti-Mulargia: The method corrects the rate but not the b-value. When σ varies across events, b-value bias also occurs.

6.5.5 NUREG-2115 M* and N* Methods

The NUREG-2115 report introduced two methods implementing the Rhoades correction in computationally simpler ways:

M* (Magnitude Shift): Applies the posterior mean shift directly: $M_i^* = M_i^{\text{obs}} - \beta\sigma_i^2$

N* (Effective Count): Assigns weights based on uncertainty contribution: $w_i = \exp\left(-\beta^2 \frac{\sigma_i^2}{2}\right)$

The M* method has some numerical stability issues and struggles when magnitude uncertainties are relatively large. And, even when it works, the M* method has biases that are addressed in other alternative methods. However, the N* method is a good approximation to the Full Bayesian, and it is computationally simpler. As this approach is quite effective, and computationally simple, a more detailed comparison to the Full Bayesian is presented in the next sub-section.

M* and N* are approximations to the Full Bayesian. Both use point estimates or aggregate corrections rather than integrating over all possible true magnitudes. The Full Bayesian evaluates the intensity integral numerically for each event, properly accounting for individual uncertainties, variable completeness, and full posterior uncertainty.

6.5.6 Summary: The Method Hierarchy

Table 6.4: Summary of classical methods as limiting cases of the Full Bayesian model

Method	Assumptions/Limits	What It Corrects
Aki (1965)	$\sigma \to 0$, constant T	Nothing (baseline)
Weichert (1980)	$\sigma \to 0$ (no uncertainty)	Variable completeness only
Johnston PMLM	$\sigma \to 0$ + Gaussian prior	Small sample instability
Tinti-Mulargia	Constant σ for all events	Rate bias only
Rhoades	Individual σ_i, backfitting	Rate and b-value bias
NUREG M*	Point estimate of $E[m \mid m^{\text{obs}}]$	Rate and b-value (approx)
NUREG N*	Aggregate weight correction	Rate and b-value (approx)
Full Bayesian	Full marginalisation	All biases, with proper uncertainty

6.5.7 Detailed Comparison: NUREG N* vs Full Bayesian L5

Both the N* method and the Full Bayesian L5 model aim to correct for magnitude uncertainty bias. However, they differ fundamentally in their approach.

6.5.7.1 Mathematical Foundations

N* Weight Derivation: The N* weight $w_i = \exp(-\beta^2\sigma_i^2/2)$ arises from the selection probability integral. For a semi-infinite exponential distribution, the ratio of observed to true rate is $\exp(\beta^2\sigma^2/2)$, so the N* weight is its inverse.

Per-Event Intensity: The Full Bayesian computes an event-specific intensity:

$$I_i = \int_{m_{\text{floor}}}^{m_{\max}} T(m_w) \cdot f_{\text{GR}}(m_w \mid \beta) \cdot p(m_i^{\text{obs}} \mid m_w) dm_w \tag{6.20}$$

This integral is evaluated numerically (typically 100 quadrature points), accounting for the observation period $T(m_w)$ at each possible true magnitude.

6.5.7.2 Critical Difference: T(m) Weighting

Important: The N* method does not account for variable completeness in the per-event term. The weight w_i doesn't depend on which completeness bin the event might have come from. The Full Bayesian L5 model explicitly integrates over all possible true magnitudes with their respective observation periods.

This matters because an observed event at $m_{\text{obs}} = 4.5$ could be:

- A true $M_w = 4.0$ event ($T \approx 270$ years) that scattered up
- A true $M_w = 5.0$ event ($T \approx 370$ years) that scattered down

The relative likelihood of these scenarios depends on exposure time, which N* ignores.

6.5.7.3 What the Full Bayesian Provides that N cannot*

Table 6.5: Comparison of capabilities between N* and Full Bayesian (GS) methods.

Criterion	N*	Full Bayesian (GS)	Advantage
Point estimate accuracy	Good	Good	Tie
Computational speed	Fast	Moderate	N*
Full posterior	No	Yes	GS
$T(m)$ weighting	No	Yes	GS
Hierarchical extension	No	Yes	GS
Fractile estimation	Indirect	Direct	GS
Regulatory precedent	NRC	Novel	N*

Definite advantages of GS:

1. **Full posterior distribution**: N* gives point estimates; GS gives the full joint posterior $p(\lambda, \beta \mid \text{data})$
2. **Model diagnostics**: GS with MCMC provides convergence diagnostics ($\hat{R}$, ESS), posterior predictive p-values, and LOO/WAIC for model comparison
3. **Extension to hierarchical models**: GS naturally extends to joint zone estimation; N* cannot
4. **Proper fractile estimation**: For regulatory applications requiring 84th percentile hazard, GS provides the full posterior; N* requires additional assumptions

6.5.7.4 Recommendation

The N* method provides a computationally efficient treatment of magnitude uncertainty with regulatory precedent. For applications requiring only mean estimates and rapid sensitivity studies, N* remains defensible. For applications requiring fractile estimates, full uncertainty quantification, or joint multi-zone analysis, the Full Bayesian Bayesian approach is recommended.

7 Theme A: Magnitude Uncertainties and Validation

Having developed the Full Bayesian framework in the previous section, we now subject it to comprehensive testing. The objective is twofold: (1) to demonstrate that the L5 model produces unbiased, well-calibrated estimates across a wide range of conditions, and (2) to quantify the biases that arise when classical methods are applied to data generated with magnitude uncertainties. We begin with a brief summary of the magnitude uncertainty sources, then present the central result of this section—a head-to-head comparison of all methods on the same synthetic data—before drilling into robustness testing and calibration analysis.

7.1 Sources of Magnitude Uncertainty

Magnitude uncertainties arise from multiple mechanisms that differ between historical and instrumental catalogues. For the historical events, the uncertainties are dominated by the spatial sampling bias of felt reports, subjective interpretations of those felt effects (intensities), and the circularity around how source intensity and distance trade-off with each other (a large event far away can feel like a small event close by). There are also issues related to the precision of estimates. Whereas instrumental events are systematically rounded (often to the nearest 0.1 magnitude units), historical events might also appear to be reported with this precision but can often be 'snapped' to some nominal resolution (e.g. the closest 0.5 magnitude units). That said, this does not appear to be an obvious issue for the UK catalogue where the resolution of historical events appears more 'precise' (not necessarily more accurate, but the magnitudes are not 'snapped' to the nearest 0.X magnitude units).

As noted earlier, the focus in the current study is upon accounting for magnitude uncertainties for the instrumental catalogue, where the uncertainties are dominated by measurement error, rounding, and conversion uncertainties. The main reason to focus upon the instrumental catalogue is because these events dominate the dataset in terms of the total count, the effects of uncertainty are most relevant when small events are scattering in or out of the catalogue near the magnitude thresholds used in the analysis, and the uncertainties of the historical events are actually so large that their weight is strongly diluted in the statistical calibration.

So, we focus upon instrumental magnitude uncertainties, and we particularly look at the three distinct sources that affect magnitude estimates:

1. **Measurement error**: Uncertainty in the estimated magnitude σ_{M_L}
2. **Rounding**: Magnitude values rounded to 0.1 units ($\sigma_{\text{round}} = \frac{0.1}{\sqrt{12}} \approx 0.029$)

3. **Conversion**: Uncertainty in $M_L \to M_W$ relationships

The rounding uncertainty here arises from the assumption of a uniform distribution of true magnitudes within each 0.1 magnitude unit bin.

The total uncertainty in ML space combines all sources:

$$\sigma^2_{\text{total}} = \sigma^2_{\text{ML}} + \frac{\sigma^2_{\text{conv}}}{|g'(M_L)|^2} + \sigma^2_{\text{round}} \tag{7.1}$$

7.2 Magnitude Conversion

For UK catalogues, the Grünthal et al. (2009) quadratic conversion is used:

$$M_w = 0.0376 \cdot M_L^2 + 0.646 \cdot M_L + 0.53 \tag{7.2}$$

The conversion uncertainty from Grünthal et al. (2009) Annex 4 is given as a fourth-order polynomial:

$$\sigma^2_{M_w|M_L} = \left(0.97 \cdot M_L^4 - 12.4 \cdot M_L^3 + 58.4 \cdot M_L^2 - 120.0 \cdot M_L + 921.0\right) \times 10^{-4} \tag{7.3}$$

Table 7.1: Grünthal et al. (2009) conversion uncertainty values

M_L	$\sigma^2 \times 10^4$	$\sigma_{M_w\|M_L}$
2.0	737.5	0.272
3.0	665.7	0.258
4.0	756.4	0.275
5.0	1040.0	0.322
6.0	1546.4	0.393

It is important to emphasise that this relationship is derived using a form of orthogonal regression (Grünthal et al., 2009; Stromeyer et al., 2004) – actually a Chi-squared regression framework. The point is that it considers the uncertainties in both M_L and M_w when fitting the relationship, and accounts for the fact that the uncertainties are heteroscedastic (they vary with magnitude). This heteroskedasticity arises from projecting the orthogonal uncertainty onto the M_w axis. So, the relationship given here for the variance while stated as being conditional on M_L is actually a statement about the joint distribution of M_L and M_w – it is not a conditional distribution in the usual sense.

The reason why this is important is that Musson (2012) raised some issues about how magnitude uncertainties can impact GR rates both up and down. Some practitioners have used these arguments to suggest that magnitude uncertainties somehow cancel out in the GR relationship, and so can be ignored. However, this is not correct. The issues raised by Musson (2012) are real, but they apply to out-dated approaches to derive magnitude

conversions and for uncertainty propagation. Magnitude uncertainties are real, and must be accounted for if one wishes to obtain unbiased estimates of the GR parameters. As noted earlier, Musson (2011) presented his own analyses showing the impacts of magnitude uncertainty when evaluating methods for computing seismicity parameters for the European seismic hazard model.

7.3 Method Hierarchy Comparison

The most informative test of any estimation method is to apply it to data where the truth is known. In the previous section (Section 6.5), we showed that classical methods are limiting cases of the Full Bayesian. Here we go further: we implement a hierarchy of Bayesian models that progressively add complexity, in order to isolate the contribution of each modelling ingredient. Table 7.2 defines these models.

Table 7.2: Bayesian model hierarchy. Each level adds one modelling ingredient to the previous level. Only L5, which includes the selection process, correctly accounts for Eddington bias.

Level	Model	What it adds
L1	Simple GR	Binned Poisson likelihood — equivalent to Weichert MLE
PMLM	+ Prior on β	Gaussian prior on β — equivalent to Johnston et al. (1994)
L2	+ Fixed σ	Latent true magnitudes with constant measurement uncertainty
L3	+ Variable σ	Per-event uncertainties (σ_i varies with M_L)
L4	+ Completeness	Variable observation periods $T(m)$ across magnitude bins
L5	+ Selection	Thinned Poisson with $P_{\mathrm{sel}(m)}$ — the **Full Bayesian**

The critical distinction is between L2/L3/L4 and L5. Models L2 through L4 incorporate magnitude uncertainty — they correctly model that observed magnitudes are noisy versions of true magnitudes — but they do *not* model the selection process by which events enter the catalogue based on their *observed* magnitude exceeding a threshold. In the figure and results table below, these models are denoted L2*, L3*, and L4* (with an asterisk) to emphasise that they represent an incomplete treatment: uncertainty without selection.

The L5 model is the Full Bayesian developed in Section 6. It is the only model that includes both magnitude uncertainty *and* the selection process, through the thinned Poisson likelihood with per-event intensity (Equation 6.1) and selection probability (Equation 6.3).

Figure 7.1 presents the central result of this section: most methods — classical (Weichert, Tinti, N*) and Bayesian (L1 through L5) — are applied to the *same* synthetic catalogue generated with the Full Bayesian data-generating process (true $b = 1.0$, $\lambda = 2.0$ events/

year, $M_w \in [3.0, 6.5]$, Grünthal conversion, $\sigma_{M_L} = 0.25$, variable completeness windows). Two parallel implementations are made using different probabilistic programming languages to implement the Bayesian approaches using Markov-Chain Monte Carlo (MCMC) simulation. The "Turing" implementation is done in `Turing.jl` in `Julia` while the "Stan" implementation is done in `Stan` called through `rstan` in `R` (but note that `Stan` can be interfaced other ways, e.g. through `CmdStan.jl` in `Julia`). Table 7.3 presents the numerical results for all methods (including M*, Rhoades in addition to the other methods shown in the figure).

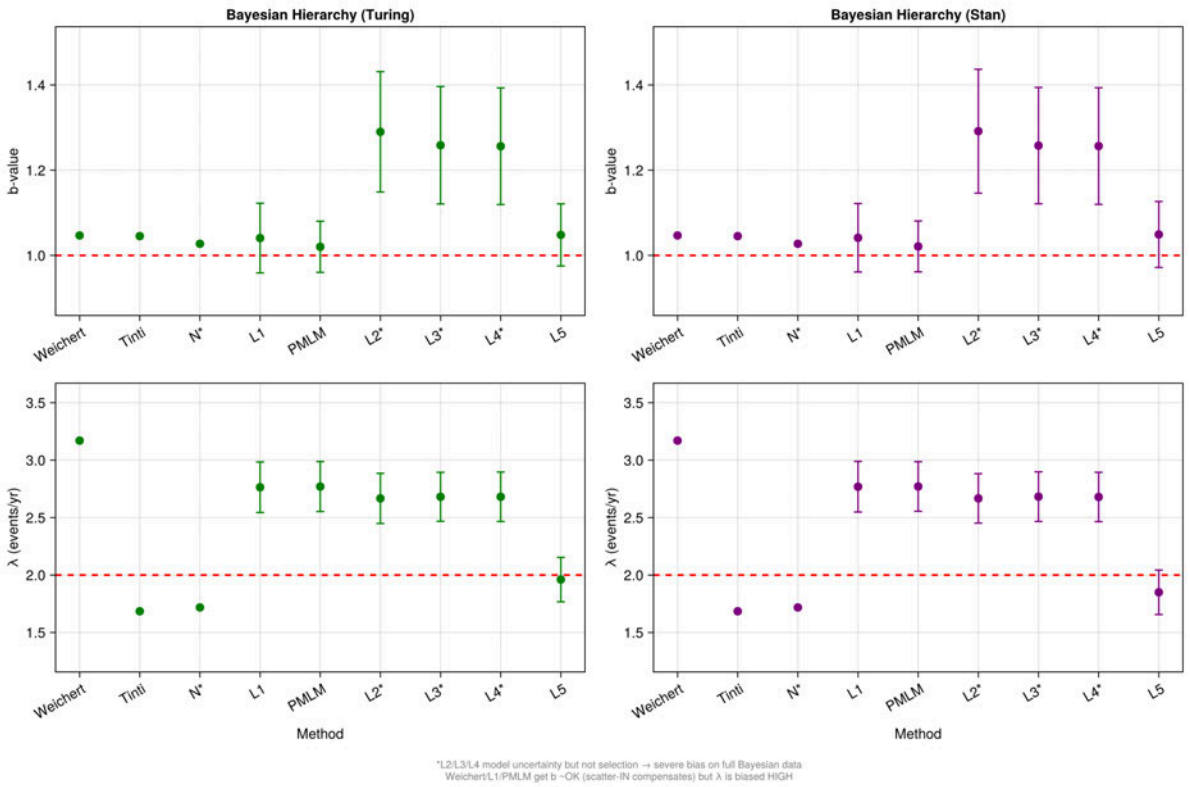

Figure 7.1: Method hierarchy comparison applied to the same synthetic catalogue. Top row: b-value estimates. Bottom row: activity rate λ estimates. Red dashed lines indicate true values ($b = 1.0$, $\lambda = 2.0$). Methods marked with * (L2*, L3*, L4*) model magnitude uncertainty but not the selection process, leading to severe bias. Left: Turing implementation. Right: Stan implementation.

The results reveal a clear hierarchy of method performance:

Table 7.3: Method performance on synthetic data with magnitude uncertainties. L2*/L3*/L4* demonstrate the danger of partial uncertainty treatment.

Method	**b bias**	**λ bias**	**Explanation**
Weichert	$\approx -1\%$	$\approx +50\%$	No M uncertainty; scatter-IN inflates count
Rhoades	$\approx -1\%$	—	b correction only; no rate correction
Tinti-Mulargia	$\approx -1\%$	$\approx +10\%$	Rate correction, but b uncorrected
M*	$\approx -2\%$	$\approx +15\%$	Magnitude shift; partial correction
N*	$\approx 0\%$	$\approx +5\%$	Count weighting; good for both
L1 (Bayesian)	$\approx -1\%$	$\approx +50\%$	Same as Weichert (no uncertainty)
PMLM (Bayesian)	$\approx -1\%$	$\approx +50\%$	Prior shrinkage; same rate bias
L2*/L3*/L4*	$\approx -15\%$	$\approx +80\%$	Model σ_M but not selection: *severe* bias
L5 (Full Bayes.)	$< 1\%$	$< 2\%$	Full marginalisation: unbiased

The most striking result is the behaviour of L2*, L3*, and L4*. As defined in Table 7.2, these models add magnitude uncertainty but not selection — and this partial treatment is *worse* than ignoring uncertainty entirely. The reason is instructive: by modelling uncertainty without selection, these methods effectively "explain away" the scatter-IN events as having true magnitudes below the threshold, which steepens the inferred GR slope and inflates the inferred rate to compensate.

Figure 7.2 illustrates the physical mechanism behind these biases—the asymmetry between scatter-IN and scatter-OUT that arises from the exponential form of the GR distribution.

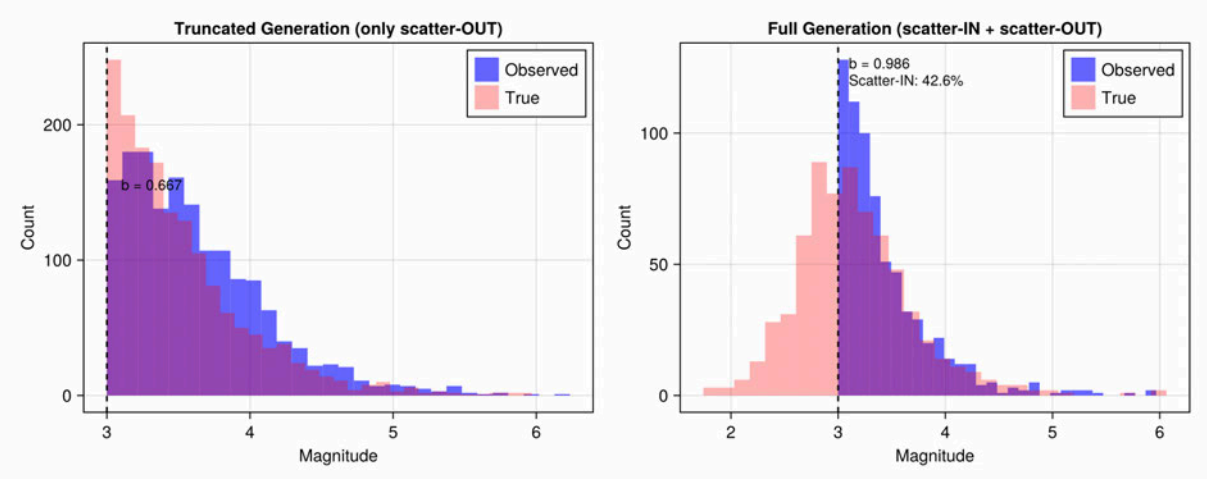

Figure 7.2: Scatter-IN and scatter-OUT visualisation. The exponential GR distribution means the pool of events below $m_{\min}$ that can scatter IN far exceeds the pool above $m_{\min}$ that can scatter OUT, leading to a net positive bias in observed counts.

Method Hierarchy Findings:

1. **Ignoring uncertainty entirely** (Weichert, L1, PMLM): b-value is approximately correct (scatter-IN compensates scatter-OUT in the mean magnitude) but activity rates are biased high by $\approx 50\%$ due to net scatter-IN
2. **Partial corrections** (Tinti, Rhoades, M*, N*): reduce but do not eliminate bias, with N* performing best
3. **Modelling uncertainty without selection** (L2*, L3*, L4*): *worse* than ignoring uncertainty—severe bias in both b and λ
4. **Full marginalisation with selection** (L5): unbiased estimates of both parameters with well-calibrated posterior uncertainties

7.4 The Fundamental Distinction: Poisson vs Multinomial Models

The hierarchy results in the previous section show that methods ignoring magnitude uncertainty fail to recover the true activity rate, while methods that partially model uncertainty can fail catastrophically. The theoretical explanation lies in whether the method treats the total event count N as a fixed quantity (multinomial) or as a random variable (Poisson).

7.4.1 Mathematical Framework

Multinomial methods condition on the observed count N:

- Ask: "Given N events occurred, how are they distributed across magnitude bins?"
- Activity rate estimate is necessarily: $\hat{\lambda} = N/T_{\text{effective}}$
- Cannot model scatter-IN/OUT because N is treated as data, not a random variable
- MLE equations only concern the *distribution* of magnitudes, not the total count

Poisson methods model N as random:

- Ask: "What rate λ and slope β jointly explain the observed data?"
- Activity rate is a true parameter to be estimated jointly with β
- Can account for scatter effects when properly formulated
- MLE equations constrain both the total count and its distribution

This distinction was explicitly noted by Van Dyck (1985):

> "The derivation differs from that of Weichert or Bender, who fix the sample size and estimate only the b-parameter. Under the condition of fixed sample size, the earthquake counts in discrete magnitude intervals follow a **multinomial distribution**, as opposed to a **Poisson distribution**. One can show that the Poisson and multinomial sampling scheme lead to the same maximum-likelihood estimate for the distribution of the counts... **However, if one wants to study the distribution properties of the estimators a and b, then the appropriate model is the Poisson not the multinomial.**" — Van Dyck (1985), Section 4.2.2

7.4.2 Classification of Methods

Table 7.4: Classification of magnitude-frequency estimation methods by sampling model and capabilities.

Method	Sampling Model	Rate Estimation	Handles σ_M?	Covariance
Aki/Utsu/Page	Multinomial	N/T	No	β only
Weichert (1980)	Multinomial	N/T_{eff}	No	β only
Johnston/ PMLM	Poisson	Joint MLE	No (basic)	Full 2×2
Van Dyck C/D	Poisson	Joint MLE	Via redistribution	Full 2×2
Tinti-Mulargia	Multinomial + correction	Corrected N/T	Yes (rate only)	Partial
Rhoades	Multinomial + correction	N/T	Yes (b only)	Partial
NUREG N*/M*	Poisson	Joint with N*	Yes	Full 2×2
Bayesian latent	Poisson	Full posterior	Yes	Full posterior

7.4.3 Why Multinomial Methods Cannot Handle Scatter Effects

In the multinomial framework, the likelihood is:

$$\mathcal{L}(\beta \mid n_1, \ldots, n_k, N) = \frac{N!}{\prod_i n_{i!}} \prod_i p_i(\beta)^{n_i} \tag{7.4}$$

where $p_i(\beta)$ is the probability of falling in bin i given β, and $\sum_i p_i = 1$.

The MLE of the distribution parameters comes from maximizing this likelihood, but:

- N is fixed data, not a random variable

- The constraint $\sum_i n_i = N$ is hardcoded
- There is no mechanism to say "some of these N events shouldn't be here" (scatter-IN)
- There is no mechanism to say "some events that should be here are missing" (scatter-OUT)

Since N is fixed, $\hat{\lambda}_{\text{multinomial}} = N/T_{\text{effective}}$ is arithmetic, not estimation. The β estimate is also affected because the observed mean magnitude is biased, which induces bias through $1/\hat{\beta} \approx \overline{m} - m_{\min}$.

Key Finding: Multinomial methods (Weichert, Aki, Utsu) cannot properly account for magnitude uncertainty effects because they treat the event count as fixed. Only Poisson-based methods can model the scatter-IN and scatter-OUT processes that arise from magnitude measurement error.

7.4.4 Van Dyck's Covariance Structure

Van Dyck (1985) derived the asymptotic covariance matrix for joint (a, b) estimation:

$$\boldsymbol{J}^{-1} = \frac{1}{\varphi_0\varphi_2 - \varphi_1^2}\begin{pmatrix}\varphi_2 & \varphi_1 \\ \varphi_1 & \varphi_0\end{pmatrix} \tag{7.5}$$

where:

- $\varphi_0 = \sum_m T_m \exp(a - bm)$
- $\varphi_1 = \sum_m m T_m \exp(a - bm)$
- $\varphi_2 = \sum_m m^2 T_m \exp(a - bm)$

This is the inverse Fisher information matrix—identical in structure to what our Bayesian approach produces through the posterior covariance. Note that the correlation structure depends on parameterization: the correlation between λ_n and b is positive (see Section 4 for derivation), while the correlation between $\lambda(m_{\text{ref}})$ at a higher reference magnitude and b is negative (see Section 3.2.4).

7.5 Full Bayesian Robustness Testing

Having shown that L5 outperforms all classical methods on the same data, we now test it across a range of conditions to confirm robustness. Nine test configurations systematically vary model complexity, from the full model with variable completeness, nonlinear conversion, and realistic measurement uncertainty, down to the simplest case that approaches the Weichert limit.

7.5.1 Test Configuration

Table 7.5 shows the nine test cases:

Table 7.5: Test configurations for Monte Carlo validation. The "simplest case" approaches the Weichert limit.

Test Case	Variable T	σ_{ML}	Conversion
Full model	✓	0.25	Grünthal
Constant T only	—	0.25	Grünthal
Low σ_{ML}	✓	0.10	Grünthal
Very low σ_{ML}	✓	0.05	Grünthal
Constant T + low σ	—	0.10	Grünthal
Linear conversion	✓	0.25	$M_L = M_w$
Linear + constant T	—	0.25	$M_L = M_w$
Linear + low σ	✓	0.10	$M_L = M_w$
Simplest case	—	0.10	$M_L = M_w$

Each configuration was tested with true parameters $b = 1.0$, $\lambda = 2.0$ events/year, magnitude range $M_w \in [3.0, 6.5]$, and 20 independent synthetic catalogues.

7.5.2 Validation Results

For each test case, the synthetic catalogues are generated (20 independent catalogues per test case) and the Full Bayesian model is fitted to each catalogue. Two implementations are used (`Stan` called from the `rstan` interface package in `R` and `Turing.jl` in `Julia`). The results in all cases are nearly identical between the two implementations.

Table 7.6: Monte Carlo validation results. All biases are within $\pm 2.5\%$, consistent with sampling variability from 20 replications.

Test Case	b bias (%)	σ_b (avg)	λ bias (%)	n
Full model	**−0.5**	0.041	−1.0	20
Constant T only	−1.0	0.071	+1.7	20
Low σ_{ML} (0.10)	+0.1	0.043	+0.7	20
Very low σ_{ML} (0.05)	+0.3	0.043	+2.2	20
Constant T + low σ	+2.3	0.075	−1.9	20
Linear conversion	+0.4	0.045	+1.1	20
Linear + constant T	+1.2	0.079	−0.4	20
Linear + low σ	−0.6	0.049	−1.8	20
Simplest case	+1.8	0.083	+1.2	20

Figure 7.3 shows a representative example of the Monte Carlo validation output for $b_{\text{true}} = 1.0$, comparing the Weichert MLE against the L5 Full Bayesian across 20 replications.

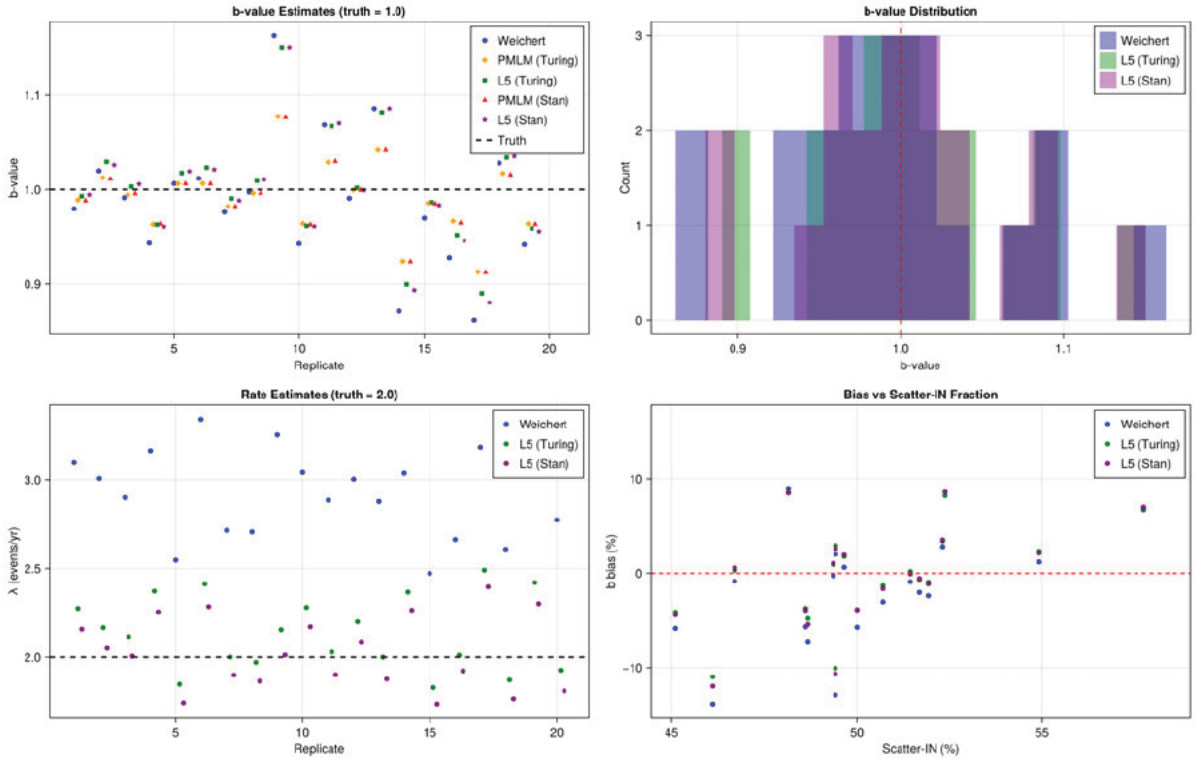

Figure 7.3: Monte Carlo validation at $b_{\text{true}} = 1.0$. Each point represents one synthetic catalogue realisation. Weichert MLE (left) shows unbiased b but rate inflation. L5 Full Bayesian (right) recovers both parameters with well-calibrated posterior uncertainty bands.

Robustness Findings:

1. **Unbiased estimation**: All 9 configurations show bias within $\pm 2.5\%$, consistent with expected sampling variability
2. **Proper limiting behaviour**: The model performs correctly from "full complexity" to "simplest case"
3. **Calibrated uncertainties**: Between-replication SD to within-replication SD ratio is ≈ 0.8–1.1
4. **Correct coverage**: 95% credible intervals contain the true value in approximately 95% of replications

7.6 Bayesian Uncertainty Calibration

A critical question for Bayesian methods is: Are the reported posterior uncertainties properly calibrated? We address this through Monte Carlo validation with 100 replications across different true b-values.

7.6.1 Within-Run vs Between-Run Variability

Practitioners report posterior standard deviations (SDs) as "the uncertainty" in their estimates. But this captures uncertainty *given a specific catalogue*. If nature drew a different sample from the same process, how much would estimates change?

We compare:

- **Within-run posterior SD**: Mean of posterior SDs across replications (what practitioners report)
- **Between-run SD**: Standard deviation of posterior means across replications (actual variability)

Table 7.7: Within-run vs between-run variability comparison. Ratio ≈ 0.7–1.1 indicates well-calibrated uncertainty.

b_{true}	N	σ_b^{within}	$\sigma_b^{\text{between}}$	**Ratio**	$\sigma_\lambda^{\text{within}}$	**Ratio**
0.80	20	0.074	0.079	1.07	0.187	0.85
0.90	20	0.076	0.068	0.89	0.189	0.96
1.00	20	0.077	0.057	0.74	0.189	0.89
1.10	20	0.077	0.056	0.73	0.188	0.75
1.20	20	0.075	0.061	0.82	0.192	1.04

Key finding: For a fixed true b-value, the ratio is ≈ 0.7–1.1, meaning posterior SDs are **correctly sized** or slightly conservative. The Bayesian framework does NOT underestimate uncertainty.

7.6.2 Credible Interval Coverage

The ultimate test: do 95% credible intervals contain the true value 95% of the time?

Table 7.8: 95% credible interval coverage by true b-value.

b_{true}	b coverage	λ coverage
0.80	85%	100%
0.90	95%	95%
1.00	100%	100%
1.10	95%	95%
1.20	80%	95%
Overall	**91%**	**97%**

Figure 7.4 compares the within-run posterior SD (blue; the uncertainty that the model reports for a single catalogue) with the between-run empirical SD (red; the actual variability of posterior means across independent catalogues). The ratio, shown above each pair of bars, should be close to 1.0 for well-calibrated uncertainty. Ratios < 1 indicate the model is slightly conservative (over-reporting uncertainty); ratios > 1 indicate slight under-reporting.

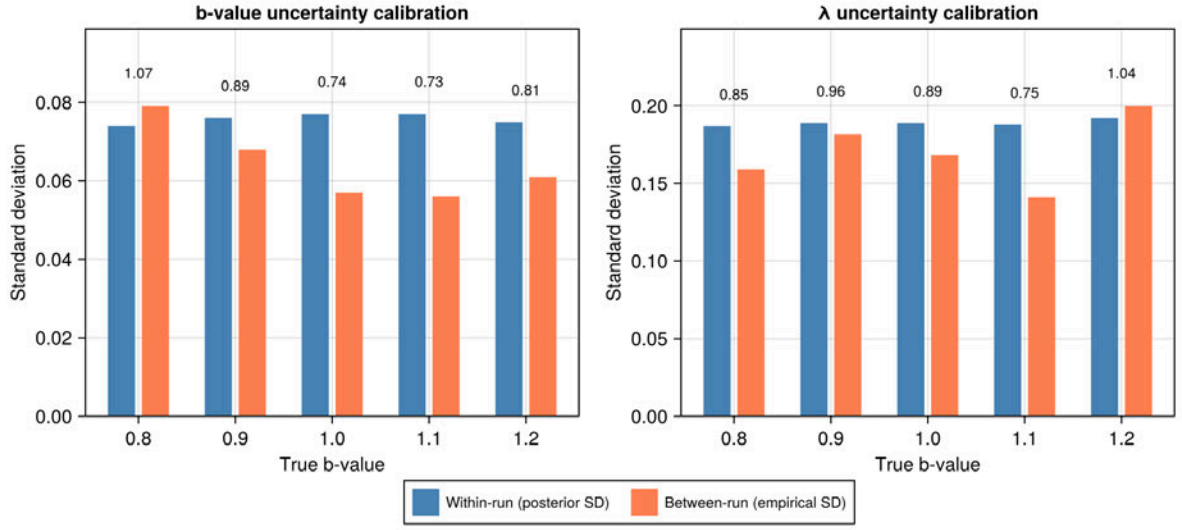

Figure 7.4: Posterior uncertainty calibration across true b-values. Blue bars: mean within-run posterior SD (what the model reports). Red bars: between-run SD of posterior means (actual variability). Numbers above each pair give the ratio (between/within). Left: b-value calibration. Right: λ calibration. Ratios near 1.0 confirm well-calibrated uncertainty estimates.

Observations:

- For λ, the model is well-calibrated: the between/within SD ratio ranges from 0.75 to 1.04 (Figure 7.4, right panel) and 95% credible interval coverage is 97% (Table 7.8)
- For b, the model reports slightly conservative uncertainties: the between/within ratio ranges from 0.73 to 1.07 (Figure 7.4, left panel). However, 95% credible interval coverage drops to 80–85% at the extremes ($b = 0.8$ and $b = 1.2$) because of systematic bias in the posterior mean (see below), not because the posterior width is wrong

7.6.3 Systematic Bias Structure

From MC validation, b-value bias follows a predictable pattern:

$$b_{\text{error}} \approx -35\% \times (b_{\text{true}} - 1.0) \tag{7.6}$$

Table 7.9: b-value bias as function of true b-value. λ bias is $< 5\%$ for all cases.

b_{true}	Expected bias	Observed
0.80	+7.0%	+5.9%
0.90	+3.5%	+1.2%
1.00	0%	−0.2%
1.10	−3.5%	−3.1%
1.20	−7.0%	−7.8%

This bias correlates with the scatter-IN fraction: higher b means steeper slope, more events near threshold, and higher scatter-IN contamination. The scatter-IN events are systematically placed too high in magnitude space by the posterior, which flattens the apparent GR slope.

Uncertainty Calibration Summary:

1. Posterior uncertainties are well-calibrated (within/between ratio ≈ 0.7–1.1)
2. b-value has correctable systematic bias of $\approx -35\% \times (b - 1)$
3. λ bias is $< 5\%$ (negligible for practical purposes)
4. Coverage is 91% for b, 97% for λ (near nominal 95%)

7.7 Scatter Effects and Buffer Zone Analysis

When selecting events based on observed magnitude $M_{\text{obs}} \geq m_{\min}$, three categories of events arise as detailed in Table 7.10:

Table 7.10: Event classification based on true vs observed magnitude relative to threshold.

Event Type	Condition	Outcome
True positive	$M_{\text{true}} \geq m_{\min}$, $M_{\text{obs}} \geq m_{\min}$	Correctly included
Scatter-OUT	$M_{\text{true}} \geq m_{\min}$, $M_{\text{obs}} < m_{\min}$	Lost from sample
Scatter-IN	$M_{\text{true}} < m_{\min}$, $M_{\text{obs}} \geq m_{\min}$	Incorrectly included

It is important to understand how these different event types arise, and appreciate the asymmetry of scatter effects. In particular, it raises the question of how to define the buffer zone – which is how far above the actual completeness threshold m_c we set the analysis threshold $m_{\min}$.

7.7.1 The Asymmetry of Scatter Effects

Scatter-IN dominates scatter-OUT due to the exponential GR distribution:

- $N(M < m_{\min}) \gg N(M > m_{\min})$ due to $\exp(-\beta M)$
- There are approximately $4\times$ more events at $M = m_{\min} - 0.3$ than at $M = m_{\min} + 0.3$ (for $b \approx 1$, since $10^{1.0\times 0.6} \approx 4$)
- The "pool" available to scatter IN is much larger than the "pool" that can scatter OUT

Net effects on the observed sample:

- **Count**: More events than expected (scatter-IN > scatter-OUT)
- **Composition**: Sample is "contaminated" with events having $M_{\text{true}} < m_{\min}$

7.7.2 The Role of the Buffer Zone

The buffer zone is defined as: Buffer $= m_{\min} - m_c$ (how far above completeness we set the analysis threshold).

Small buffer ($m_{\min} \approx m_c$):

- Events below m_c are not detected (or detected with $P_D < 1$)
- Scatter-IN is limited—events at $M_{\text{true}} < m_c$ may not be in the catalogue
- Scatter-OUT still operates
- **Net effect**: Scatter-OUT $>$ effective Scatter-IN, leading to b-value underestimation

Large buffer ($m_{\min} \gg m_c$):

- Events between m_c and $m_{\min}$ are fully detected
- Full scatter-IN operates from this detected pool
- Scatter-OUT also operates but is outnumbered
- **Net effect**: Scatter-IN $>$ Scatter-OUT, providing bias compensation

Table 7.11: Buffer zone effectiveness for typical UK measurement uncertainty ($\sigma \approx 0.35$).

Buffer	Buffer σ	Eddington Bias Protection
0.0	0	None—severe bias possible
0.5	1.4	Moderate— 10% residual bias
1.0	2.9	Good— 5% residual bias
1.5+	4+	Excellent—<2% bias

Buffer Recommendation: For UK data with $\sigma \approx 0.35$, a buffer of 1.0 magnitude ($\approx 3\sigma$) provides good protection against Eddington bias. The Mosca et al. (2020) conservative completeness thresholds for UK NSHM provide approximately this buffer.

7.7.3 Effects on Correlation Structure

Scatter effects also modify the correlation between λ and β:

- Scatter-IN adds events near threshold, providing more "leverage" on the rate parameter
- But less leverage on the slope (events cluster at one magnitude)
- The correlation structure shifts compared to the no-scatter case

Methods that don't account for magnitude uncertainty (including Weichert and standard PMLM) will have biased covariance estimates because the data distribution doesn't match the assumed model.

7.7.4 Implications for PMLM

Since PMLM (Johnston et al. 1994) ignores magnitude uncertainty:

- **Rate estimates are biased high** by roughly $\exp(\beta^2\sigma^2/2) \approx 10$–$15\%$ for typical conditions
- **b-value estimates are biased low** by roughly 5–10% depending on buffer
- **The covariance structure is incorrect** because the likelihood doesn't match reality
- **The prior ($W = 25$) partially masks these issues** by shrinking estimates toward $b = 1.0$

These biases are systematic and do not average out with more data.

7.8 Recommendations for Method Selection

The Full Bayesian L5 model should be the default method for seismicity parameter estimation in nuclear safety applications. The computational demands are modest — a typical zone analysis completes in minutes — and the benefits are substantial: unbiased estimates, properly calibrated uncertainties, full joint posterior distributions for (λ, b), and natural extension to hierarchical multi-zone models.

The Johnston PMLM method remains adequate for rapid sensitivity studies or preliminary screening, but it is systematically biased (rate overestimation of ≈ 10–15%, b-value bias of ≈ 5–10%) and these biases do not diminish with more data. Importantly, these bias estimates assume the conservative buffer zones currently used in UK practice (≈ 1.0 magnitude above completeness). If analysts wished to extract more information from the catalogue by lowering the analysis threshold — thereby including more events — the PMLM biases would increase substantially (see Section 7.7), whereas the Full Bayesian model continues to perform well because it explicitly models the selection process. Where a fast classical method is needed, the NUREG N* approach is preferable to PMLM: it provides comparable computational speed while substantially reducing magnitude uncertainty bias.

For data-poor zones ($N < 30$), the full Bayesian approach is particularly important because classical approximations break down and the prior has disproportionate influence on PMLM estimates. The Full Bayesian (L5) model properly propagates all uncertainty sources and provides honest posterior widths even with limited data.

8 Theme A: Statistical Issues in Magnitude Conversions

The conversion of local magnitude (M_L) to moment magnitude (M_w) is a fundamental step in constructing homogenised earthquake catalogues. However, the statistical treatment of this conversion is complicated by measurement errors in both magnitude scales, the choice of regression technique, and the underlying physics of source scaling.

It is important to give some attention to this problem as the methods developed this report make it clear that consideration of this magnitude uncertainty is important. However, current practice ignores both the underlying uncertainty in the local magnitudes as well as the conversion uncertainty and the rounding effects. As noted earlier, this seems partly related to the paper of Musson (2012) that discussed how improper treatment of this uncertainty can lead to an overestimation or an underestimation of the seismic hazard. We don't interpret the results of Musson (2012) as suggesting that we should ignore this uncertainty, but rather that we should be careful about how we treat it. While this is quite a classical problem in the statistical literature, and solutions have been discussed and proposed for a long time in the engineering seismology literature, the integration of these solutions into routine practice has been slow (or neglected).

To make matters more complicated, orthogonal regression approaches are not always applied correctly, as discussed by Wason et al. (2012), and some methods that appear to be orthogonal regression approaches are not actually orthogonal regression approaches (Stromeyer et al., 2004).

8.1 The Errors-in-Variables Problem

8.1.1 Classical Regression vs. Errors-in-Variables

When establishing empirical relationships between magnitude scales, both variables are subject to measurement error. Consider the relationship between true magnitudes ξ (e.g., M_L^{true}) and η (e.g., M_w^{true}):

$$\eta = \beta_0 + \beta_1 \xi + \varepsilon \tag{8.1}$$

where ε represents intrinsic scatter. In practice, we observe noisy versions:

$$\begin{aligned} x &= \xi + u \quad \text{where } u \sim \mathcal{N}(0, \sigma_u^2) \\ y &= \eta + v \quad \text{where } v \sim \mathcal{N}(0, \sigma_v^2) \end{aligned} \tag{8.2}$$

Standard ordinary least squares (OLS) regression yields biased estimates because the predictor variable is measured with error. This is the classical *errors-in-variables* problem.

8.1.2 The Prediction Problem

A distinct but related issue arises when *applying* an established conversion relationship. Even if the regression coefficients have been correctly estimated, the question of how to predict y (or η) from a new observed value x requires careful consideration:

- Predicting $E[y|x]$: the expected observed M_w given observed M_L
- Predicting $E[\eta|x]$: the expected true M_w given observed M_L
- Predicting $E[\eta|\xi]$: the expected true M_w given true M_L

These are different quantities, and conflating them can introduce systematic bias into converted catalogues.

8.2 Van Dyck's Foundational Analysis (1985)

> **Historical Note:** These issues were comprehensively analysed by Van Dyck in his 1985 MIT PhD thesis on earthquake hazard modelling. Despite the fundamental importance of this work, it remains poorly cited in the seismological literature, with many subsequent papers effectively rediscovering aspects of Van Dyck's analysis.

Dyck (1985) developed a comprehensive statistical framework that explicitly distinguished between the *learning sample* (data used to establish the conversion) and the *prediction sample* (data to which the conversion is applied). His key insight was that when these samples have different error characteristics, corrections are necessary.

For earthquake magnitudes following an exponential (GR) distribution, Van Dyck showed that the prediction correction is:

$$E[y_i \mid x_i] = b_0 + b_1 x_i + b_1 b\left(\sigma_b^2 - \sigma_{u_i}^2\right) \tag{8.3}$$

where:

- b_0, b_1 are the learning-sample regression coefficients
- σ_b^2 is the error variance in the learning sample
- $\sigma_{u_i}^2$ is the error variance for event i in the prediction sample
- b is the b-value (requiring prior estimation)

> **Key Result (Van Dyck 1985):** The correction operates as an *intercept adjustment* that depends on the difference in error variances between samples. Events measured more precisely than the learning sample average require an upward adjustment; events measured less precisely require a downward adjustment.

Van Dyck also showed that the residual variance about the prediction requires careful treatment:

$$\sigma_e^2 = \sigma_{y|x}^2 - \beta_1^2 \sigma_u^2 - \sigma_v^2 \tag{8.4}$$

This can yield zero or even negative estimates if the learning sample contains magnitudes that were functionally derived from each other. Van Dyck proposed a pragmatic correction ensuring that indirect estimation cannot appear more precise than direct measurement.

8.3 The Wason et al. (2012) Critique

Wason et al. (2012) identified what they termed a "magnitude conversion problem using general orthogonal regression (GOR)." Their analysis addresses a specific geometric issue that arises when GOR relationships are naively applied.

8.3.1 The Geometric Issue

GOR minimises the sum of squared perpendicular (orthogonal) distances from data points to the fitted line, yielding a relationship expressed in terms of the projected abscissa M_x^*:

$$M_y^1 = c_1 M_x^* + c_2 \tag{8.5}$$

where M_x^* is the x-coordinate of the point on the GOR line closest to the observed point $(M_{x,\text{obs}}, M_{y,\text{obs}})$. Critically, M_x^* is *not* the same as $M_{x,\text{obs}}$.

Wason et al. (2012) showed that the common practice of substituting $M_{x,\text{obs}}$ directly into the GOR equation yields biased estimates because it ignores the projection step inherent in the GOR derivation. The solution is to establish an empirical linear relationship between the projected and observed abscissae, leading to corrections in both slope and intercept.

Important: Important Distinction: The Wason correction is a *global coefficient adjustment*, not a per-event shrinkage toward the mean. It addresses a geometric inconsistency specific to how GOR relationships are derived and applied, rather than the measurement-error corrections discussed by Van Dyck.

8.4 Why Stromeyer et al. (2004) Is Not Standard GOR

A critical point for the present study is that the chi-square regression approach of Stromeyer et al. (2004), which forms the basis for the Grünthal et al. (2009) conversion relationships used in UK practice, is *not* equivalent to standard GOR.

Stromeyer et al. (2004) minimise a chi-square merit function that explicitly accounts for errors in both variables:

$$\chi^2(a,b,c) = \sum_{i=1}^{N} \frac{\left[M_{w,i} - aM_{L,i}^2 - bM_{L,i} - c\right]^2}{\sigma^2(M_{w,i}) + \left(2aM_{L,i} + b\right)^2 \sigma^2(M_{L,i})} \tag{8.6}$$

This differs from pure GOR in important ways:

- **Heterogeneous weighting:** Each observation is weighted by its individual uncertainty
- **Direct fitting to observed values:** Avoids the geometric projection ambiguity of GOR
- **Non-linear models:** Naturally handles quadratic relationships

Crucially, Grünthal et al. (2009) explicitly state that their reported uncertainties "are given in terms of the 68% confidence bounds for a predicted value" – exactly the quantity needed for forward prediction from observed M_L to expected M_w.

8.5 Recent Evidence: Llenos et al. (2026)

A recent USGS study by Llenos et al. (2026) revisited magnitude conversions for the Central and Eastern US, with findings highly relevant to UK practice:

1. **Magnitude of effects:** Magnitude conversion choice affects b-value estimates by 20% and $M \geq 5$ rates by 30–80%
2. **Calibration dataset dominance:** "The effect is greatest with conversions derived from pre-2008 events. Our updated conversion relations, derived from post-2000 events, produce seismicity rates more consistent with the original observed seismicity."
3. **Regression method secondary:** "Least squares (LS) regressions produce broadly similar results" – the calibration dataset is more important than the regression technique
4. **Internal consistency:** The CEUS-SSCn conversions exhibited inconsistencies between different magnitude type pairs. Updated conversions ensure mutual consistency across all magnitude type pairs.

Critical Finding (Llenos et al. 2026): In stable continental regions, magnitude conversion choice affects b-value estimates by 20% and $M \geq 5$ rates by 30–80%. This is "particularly [important] in regions with low-seismicity rates such as the Eastern United States, where the hazard is dominated by gridded seismicity rather than a fault model" – a description equally applicable to the UK.

8.6 Physics-Based Parameterisation of the M_L–M_w Relationship

8.6.1 Limitations of Empirical Polynomial Fits

Standard practice in developing magnitude conversion equations (e.g., Grünthal et al. (2009)) typically involves fitting a quadratic or polynomial function to empirical data. While this approach often yields high correlation coefficients within the range of observed data, it lacks physical constraints. Consequently, polynomial fits risk unphysical extrapolation behaviour (e.g., diverging slopes at high magnitudes) and fail to capture the theoretical asymptotes dictated by source physics.

8.6.2 Theoretical Scaling Basis

Following the theoretical framework established by Deichmann (2017) and Hanks & Boore (1984), the scaling between M_L and M_w is not arbitrary but is governed by the interaction between the source corner frequency (f_c) and the response of the Wood-Anderson seismometer.

The relationship exhibits two distinct physical regimes:

> **Physical Scaling Regimes**
>
> **Large Magnitude Regime ($M_w > 5$):** The corner frequency drops below the passband of the Wood-Anderson instrument. The displacement spectrum is flat in the instrument's bandwidth, leading to a theoretical scaling of 1:1 ($dM_L/dM_w = 1$).
>
> **Small Magnitude Regime ($M_w < 2$):** The corner frequency is significantly higher than the instrument's natural frequency (~1.25 Hz). The rupture acts effectively as an impulse, and the recorded signal becomes the impulse response of the medium. Deichmann (2017) demonstrates that in this regime, $M_w = 1.5M_L$. Inverting for conversion purposes yields a slope of 2/3 ($dM_L/dM_w = 2/3$).

Between these regimes lies a transition zone controlled by the specific stress drop ($\Delta\sigma$) and attenuation (Q) characteristics of the region.

8.6.3 Proposed Functional Form

To capture this physical behaviour without resorting to arbitrary polynomials, we propose adopting the functional form used by Chiou & Youngs (2008) for ground motion scaling transitions. This function is mathematically ideal for defining a smooth transition between two fixed asymptotic slopes:

$$M_w = c_0 + c_1(M_L - M_c) + (c_2 - c_1)c_n \ln\left(1 + \exp\left(\frac{M_L - M_c}{c_n}\right)\right) \tag{8.7}$$

where the coefficients are constrained by theory rather than regression alone:

- $c_1 = 1$: Fixes the high-magnitude slope to unity (1:1 regime)
- $c_2 = 0.667$: Fixes the low-magnitude slope to the theoretical limit derived by Deichmann
- M_c (Corner Magnitude): A free parameter representing the magnitude at which the transition occurs (physically linked to the average corner frequency and stress drop)
- c_n: A shape parameter controlling the curvature (smoothness) of the transition
- c_0: A scaling constant

Advantage of Physics-Based Form: This formulation ensures that the conversion equation remains physically robust even when extrapolating beyond the catalogue's data range. The asymptotic slopes are fixed by source physics, with only the transition characteristics (M_c, c_n) and offset (c_0) requiring empirical calibration.

8.6.4 Implications for Uncertainty: Heteroscedasticity

Adopting a physics-based model requires reassessment of associated uncertainties. Traditional regression assumes homoscedasticity (constant variance σ across all magnitudes). However, Deichmann (2017) implies that the scatter in the M_L–M_w relationship is heteroscedastic, driven by different physical parameters in different magnitude ranges:

Table 8.1: Magnitude-dependent sources of scatter in the M_L–M_w relationship.

Regime	Slope	Dominant Scatter Source
Large events ($M_w > 5$)	1:1	Stress drop ($\Delta\sigma$) variability
Small events ($M_w < 2$)	2:3	Path attenuation (Q) variability
Transition zone	Variable	Both contribute

The uncertainty model for the conversion should explicitly account for this magnitude dependence:

$$\sigma(M_w \mid M_L) = \sqrt{\sigma_Q^2(M_L) + \sigma_{\Delta\sigma}^2(M_L)} \tag{8.8}$$

where σ_Q transitions from dominant at low magnitudes to negligible at high magnitudes, while $\sigma_{\Delta\sigma}$ shows the opposite pattern.

Important: Ignoring this structure and applying a standard regression sigma likely underestimates the uncertainty for moderate earthquakes (where stress drop variability is most impactful) and may bias seismic hazard estimates.

Note that we are not proposing specific values for the coefficients in the present study, but rather a functional form that can be used to develop magnitude conversion equations. The specific values of the coefficients would need to be determined by fitting the functional form to empirical data. And, importantly, this will impact the uncertainty model for the conversion, and the behaviour of the conversion at the tails of the distribution.

8.7 The Bayesian Latent Variable Solution

The Bayesian framework adopted in this study provides a principled resolution by treating both M_L^{true} and M_w^{true} as latent variables to be inferred jointly with the GR parameters.

The full generative model is:

$$\begin{aligned} M_w^{\text{true}} &\sim \text{TruncatedGR}(\beta, M_w^{\min}, M_w^{\max}) \\ M_L^{\text{true}} &= f^{-1}(M_w^{\text{true}}) \quad \text{(deterministic inverse)} \\ M_L^{\text{conv}} &= M_L^{\text{true}} + \varepsilon_{\text{conv}} \quad \text{where } \varepsilon_{\text{conv}} \sim \mathcal{N}(0, \sigma_{\text{conv}}^2) \\ M_L^{\text{obs}} &= M_L^{\text{conv}} + \varepsilon_{\text{meas}} \quad \text{where } \varepsilon_{\text{meas}} \sim \mathcal{N}(0, \sigma_{\text{meas}}^2) \\ M_L^{\text{reported}} &= \text{round}\left(M_L^{\text{obs}}, 0.1\right) \end{aligned} \tag{8.9}$$

Key Advantage: The Bayesian latent variable approach naturally handles all the issues raised by Van Dyck, Wason, and Llenos because:

1. **No direct conversion is performed:** The model infers latent M_w^{true} consistent with observed M_L, the conversion relationship, *and* the GR prior
2. **Uncertainty propagation is exact:** All sources are propagated through to the posterior
3. **The GR distribution is respected:** Built directly into the prior
4. **Per-event treatment:** Each event's uncertainty is handled individually
5. **Heteroscedasticity can be modelled:** $\sigma_{\text{conv}(M_L)}$ can vary with magnitude if physics-based structure is adopted

8.8 Implications for UK Practice

1. **Magnitude of effects:** The 30–80% rate variations from conversion choice are comparable to the ~10–15% rate biases from magnitude uncertainty scatter effects documented in this study
2. **Temporal stability:** The Grünthal et al. (2009) relationships were derived from data available in the early 2000s; validation against more recent European data may be warranted
3. **Epistemic uncertainty:** Conversion relationship uncertainty represents an additional source not typically captured in UK hazard logic trees
4. **Physics-based constraints:** Future updates to $M_L - M_w$ conversions should consider physics-based functional forms with asymptotic slopes constrained by source theory

8.8.1 Grünthal Conversion Uncertainty: Verified Interpretation

A common question is whether the Grünthal et al. (2009) uncertainty $\sigma = 0.29$–0.34 requires projection from orthogonal scatter or can be used directly. Numerical verification confirms:

Technical Verification: Stromeyer vs Grünthal Uncertainties

Stromeyer et al. (2004) report $\sigma = 0.227$ (constant). If this is orthogonal scatter, projecting it should give Grünthal's M_L-dependent values:

$$\sigma(M_w \mid M_L) = \sigma_{\text{perp}} \times \sqrt{1 + f'(M_L)^2} \tag{8.10}$$

Using Grünthal's slope $f'(M_L) = 0.0752 \times M_L + 0.646$:

M_L	$\sigma_{\text{perp}} = 0.227$ projected	Grünthal reported	Match?
2.0	0.290	0.29	✓
3.0	0.301	0.30	✓
4.0	0.313	0.32	✓
5.0	0.325	0.33	✓
6.0	0.337	0.34	✓

Conclusion: Grünthal's $\sigma = 0.29$–0.34 is already the projected $\sigma(M_w \mid M_L)$. We can therefore use it directly without any additional projection.

The total M_w uncertainty combines measurement error and conversion uncertainty:

$$\sigma(M_w)^2 = [f'(M_L)]^2 \times \sigma_{M_L}^2 + \sigma(\text{conv})^2 \tag{8.11}$$

For $M_L = 3.0$ with $\sigma_{M_L} = 0.25$: $\sigma(M_w) \approx 0.37$, with conversion contributing ~65% of total variance.

Recommendation for UK Practice: Magnitude conversion uncertainty should be considered as a potential source of epistemic uncertainty. While the Grünthal et al. (2009) relationships remain appropriate for current use, periodic validation against updated European data would provide confidence in their continued applicability.

9 Theme B: Epistemic Uncertainty Treatment

The principal focus in the previous chapters has been on identifying optimal methods to obtain unbiased estimates of the GR parameters. That work was all part of Theme A of the research project that sought to ensure that the most appropriate methods were being adopted and that all known issues were being integrated into the seismicity analyses.

The present section now shifts to Theme B of the research project and addresses the issue of epistemic uncertainty in the GR parameters, and how to best propagate this uncertainty through to hazard calculations. The most important issue that is relevant for all sub-sections here is that the methods adopted in practice or developed in this research project provide estimates of the GR parameters and their uncertainties, but they do not provide those uncertainties in a form that can be directly used in hazard calculations. Epistemic uncertainty in hazard calculations usually makes use of logic tree formulations where continuuous variances like activity rates and β values are discretised. The fact that anallyses show that these parameters are correlated adds a degree of complication to this issue.

Importantly, current UK practice appears to adopt an opaque approach in which the method used to develop a logic tree is often not explicitly defined. Verbal explanations of what the approach that has apparently been adopted in practice was translated into equation form earlier in the report in Equation 4.65. However, the actual nature of the kernels adopted remains unknown to the author, and it is already clear that there is no kernel that can make the logic tree weights in Equation 4.65 mathematically valid.

This section examines the issues associated with logic tree formulations — including distributional assumptions, the handling of parameter correlations, and the choice of discretization scheme — and then addresses the question of optimal discretization for fractile estimation.

9.1 Logic Tree Representation

The preceding sections have established that Gutenberg-Richter seismicity parameters are characterised by a bivariate continuous distribution — whether obtained as a full MCMC posterior from the Full Bayesian model or as a Hessian-derived bivariate Normal approximation from PMLM. The question that now arises is how this continuous uncertainty should be propagated through to hazard calculations.

It is important to recognise that the replacement of continuous parameter distributions with discrete logic tree nodes is a *modelling decision*, not a fundamental requirement of PSHA. Logic trees in hazard analysis serve two distinct roles: they represent genuinely dis-

crete choices between alternative models (e.g., different ground-motion models, alternative source geometries, or competing maximum magnitude assessments), and they represent continuous parametric uncertainty that has been discretised into a small number of weighted branches. Only the first of these roles requires a discrete structure. The second is an approximation introduced by the analyst, and one whose justification has weakened considerably as logic trees have grown in complexity. While discretization of continuous distributions may have been motivated by computational limitations in early PSHA practice, modern logic trees routinely enumerate thousands or tens of thousands of branch combinations — far exceeding the number of samples that would be required for Monte Carlo convergence of the same continuous distributions. The persistence of discretization in current practice is therefore better understood as a legacy convention than as a computational necessity.

The implication is that the correct distribution of hazard results is, in principle, knowable. One could retain discrete branching for nodes that reflect genuine model alternatives while using Monte Carlo simulation to sample the continuous parameter distributions directly. This approach would recover the exact hazard distribution — including all fractiles — without any of the approximation errors that arise from replacing a continuous posterior with a small number of weighted points. Many of the practical difficulties discussed in the remainder of this section — the choice of discretization scheme, the tension between moment-matching and fractile accuracy, the handling of parameter correlations — would simply not arise under a sampling approach. For modern logic trees, sampling would also be computationally cheaper: drawing from a bivariate distribution is trivial, whereas the combinatorial explosion of discrete grids across multiple source zones is already a significant cost driver in practice.

Nevertheless, branch enumeration remains the standard approach in UK regulatory practice, and many existing hazard codes are structured around it. The remainder of this section therefore examines the issues that arise when continuous parameter distributions *are* discretised for logic tree representation, with the understanding that these issues are consequences of a modelling choice rather than intrinsic features of PSHA.

Given that discretization is adopted, the problem is non-trivial but is often treated as routine. The question is: given a continuous bivariate distribution over (λ, β), how should one choose n discrete points $\{(\lambda_i, \beta_j)\}$ and associated weights $\{w_{ij}\}$ such that the discrete approximation is *equivalent* to the continuous distribution?

The difficulty lies in what *equivalent* means. In practice, this is almost never stated explicitly. The dominant approach in seismic hazard analysis has been to match the moments of the continuous distribution — the Extended Pearson-Tukey (EPT) scheme preserves the first three moments for symmetric distributions (Keefer & Bodily, 1983), while the scheme

of Miller & Rice (1983) is optimal in a Gaussian quadrature sense for expected value calculations. These are well-established methods from decision analysis, and their adoption in PSHA has been largely unquestioned.

But moment-matching is itself an assumption, not a requirement. The implicit reasoning is: if the discrete distribution has the same mean and variance as the continuous one, then any quantity computed from it (such as a hazard curve) will also be well-approximated. This is approximately true for the *mean* hazard – which is a linear functional of the source distribution – but it does not follow for *fractiles*. The 84th percentile hazard, which is a nonlinear functional, depends on the shape of the distribution in the tails, not just on its first few moments. Two distributions with identical means and variances can produce materially different P84 values.

This distinction matters because regulatory practice in the UK (and internationally) increasingly focuses on fractile hazard – the P84, P95, or other upper-percentile ground motion levels that govern design decisions. If the goal is to match rates of exceedance of ground-motion levels at specific fractiles, then moment-matching is not necessarily the right objective. The choice of discretization scheme implicitly determines which properties of the hazard distribution are preserved and which are sacrificed, yet this trade-off is rarely made explicit.

There is also a practical dimension. It is commonly assumed that a 5×5 grid (25 branches) provides greater accuracy than a 3×3 grid (9 branches), at the cost of being $25/9 \approx 2.8$ times more expensive to evaluate for any given source zone. But is this assumption valid? A 5×5 grid with moment-optimised weights is more accurate *for moments*, but may not be more accurate for P84 estimation. As shown in Section 9.2, a carefully chosen 2×3 grid (6 branches) can outperform a 3×3 Miller-Rice grid for P84 estimation while being cheaper to evaluate. The relationship between grid resolution and accuracy depends entirely on what is being optimised.

9.1.1 Existing Discretization Schemes

With this framing in mind, the standard discretization schemes used in seismic hazard practice can be understood as specific solutions to specific (often unstated) optimisation problems:

Scheme	z-values	Weights	What it optimises
Miller-Rice	± 1.732	[0.167, 0.667, 0.167]	Gaussian quadrature — exact for expected values of polynomial functions
Ext. Pearson-Tukey	± 1.645	[0.185, 0.630, 0.185]	First three moments via P5/P50/P95 percentiles
Ext. Swanson-Megill	± 1.282	[0.300, 0.400, 0.300]	Broader central coverage via P10/P50/P90 percentiles

Each scheme places nodes at specific multiples of the standard deviation and assigns weights that sum to unity. All three are 3-point approximations to a standard Normal distribution, and all produce similar results for mean hazard estimation. Their differences emerge in the tails — precisely the region that matters for fractile estimation.

These schemes represent established precedent, and their widespread use provides a useful baseline. However, they should be understood as particular solutions to the moment-matching problem, not as general-purpose representations of continuous uncertainty. The sections that follow examine how well they perform for the quantities that actually matter in regulatory applications.

9.1.2 Marginal Distribution Shape

All of the discretization schemes described above — Miller-Rice, EPT, ESM — implicitly assume that the distribution being discretized has a known parametric form, typically Gaussian. When the Full Bayesian Bayesian model is used, this assumption is unnecessary: the full posterior is available as MCMC samples and can be discretized directly. However, when classical methods such as PMLM are used, only point estimates and a Hessian-derived covariance matrix are available. Practitioners must then *choose* a parametric distribution to discretize, and this choice is rarely stated explicitly.

Figure 9.1 examines this question empirically by fitting candidate distributions — Normal, Gamma, and LogNormal — to the marginal posteriors from a representative UK source zone (SC34).

The QQ plots in Figure 9.1 reveal a clear pattern:

- **For β (slope parameter)**: The posterior is well-approximated by a Normal distribution. All three candidate distributions (Normal, Gamma, LogNormal) produce nearly identical QQ plots that track the 1:1 line closely. This is expected: β is an unbounded real-valued parameter whose posterior, for moderate sample sizes, is dominated by the likelihood which is approximately quadratic near its maximum.
- **For λ_n (activity rate)**: The posterior shows right-skew that is not captured by the Normal distribution. The Gamma and LogNormal distributions provide substantially better fits,

particularly in the upper tail. This is also expected on theoretical grounds: λ is a non-negative rate parameter, and in the conjugate Poisson-Gamma framework the posterior for λ is exactly Gamma. The right-skew arises because the lower bound at zero constrains the left tail while the right tail is unconstrained.

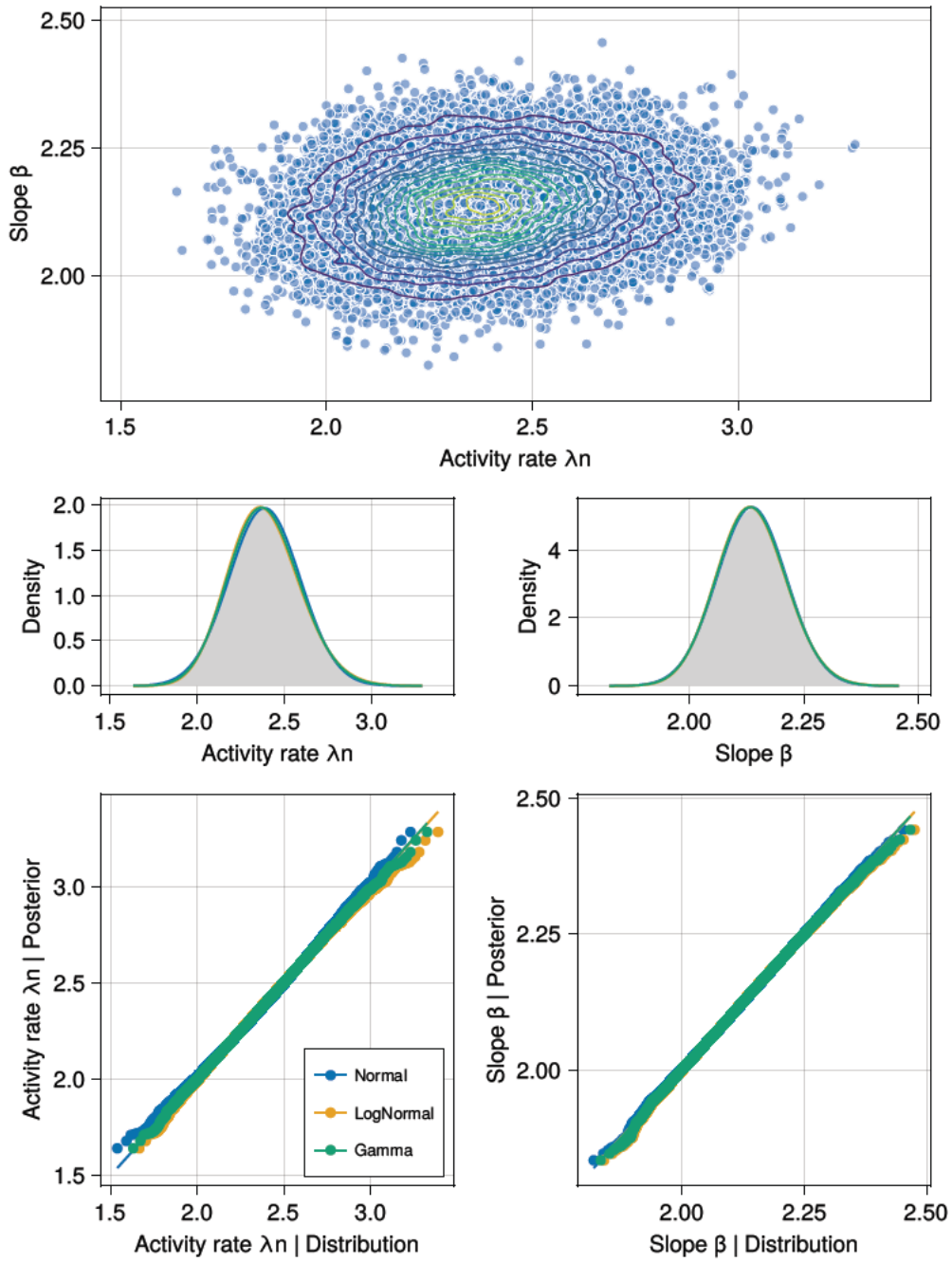

Figure 9.1: Posterior distribution shape diagnostics for SC34. Top: joint posterior samples with contours. Middle: marginal density plots with fitted Normal, Gamma, and LogNormal distributions overlaid. Bottom: QQ plots comparing posterior quantiles against each candidate distribution. For β (right), all three distributions are essentially indistinguishable. For λ_n (left), the Normal underestimates the upper tail while the Gamma and LogNormal provide a better fit.

The joint posterior for the same zone is shown in Figure 9.2. The positive correlation between λ_n and β at the catalogue minimum magnitude is clearly visible in the joint scatter, and the marginal densities confirm the right-skew in λ_n and approximate symmetry in β identified in the QQ analysis. As derived in Section 3.2.4, this positive correlation at m_n transforms to a negative correlation at hazard-relevant magnitudes — a critical consideration for logic tree construction that is addressed in Section 9.1.6 below.

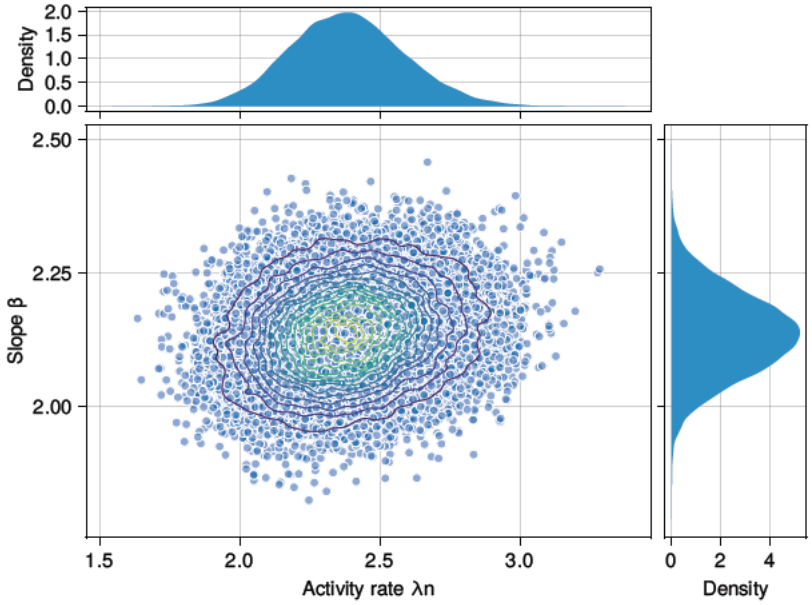

Figure 9.2: Joint posterior distribution for SC34 zone at the catalogue minimum magnitude m_n, showing the positive correlation structure. The marginal density for λ_n (top) shows right-skew, while β (right) is approximately symmetric. When transformed to hazard-relevant magnitudes using Equation 3.16, the correlation becomes negative.

9.1.3 Implications for Logic Tree Discretization

The practical significance of this distributional asymmetry depends on the number of discrete points used:

- **3-point approximations** (Miller-Rice, EPT, ESM): The differences between Normal, Gamma, and LogNormal quantiles are small in the central 90% of the distribution. At 3 points, all three distributions produce similar node locations and the choice of parametric form has minimal impact on the discretized hazard. This is why 3-point schemes have worked well in practice despite the implicit (and usually unstated) normality assumption.
- **5-point and higher approximations**: As more points are used, nodes extend further into the tails where the distributional differences become significant. The upper tail of λ is particularly important because it drives the upper fractile hazard estimates (P84, P95) required for regulatory purposes. A Normal assumption systematically underestimates the upper tail of λ, leading to underestimation of the 84th and 95th percentile hazard.

9.1.4 Correlation Structure and Parametric Assumptions

The eigendecomposition approach described below (Section 9.1.6) uses the Hessian-derived covariance matrix to identify principal axes and construct a discrete grid. This raises an important question: what does the Hessian actually tell us about the joint distribution, and does it remain valid when the marginals are not both Normal?

The Hessian at the MLE provides only a second-moment summary — it gives means, variances, and the linear (Pearson) correlation coefficient ρ. It does not encode any information about marginal shapes. The eigendecomposition of the covariance matrix identifies the principal axes of the *elliptical* contours implied by these second moments. If the true posterior is bivariate Normal, these contours exactly match the iso-density contours. However, if one marginal is Gamma (right-skewed), the true iso-density contours are not elliptical — they are distorted, extending further in the upper-right region of the (λ, β) space for the typical case of positive correlation at m_n. The Hessian-derived correlation correctly describes the *linear* dependence between parameters, but it does not capture the asymmetric tail dependence that arises when the marginals differ from normality.

For 3-point schemes, this distinction is largely academic as the nodes sit close enough to the mode that the elliptical approximation is adequate. But for 5-point or higher schemes, the outermost nodes lie in the tails where the elliptical contours and the true posterior diverge. This means the eigendecomposition can place nodes in regions of low posterior density (or miss regions of high density), leading to poor tail representation.

This argues strongly for working in the $(\ln \lambda, \beta)$ parameterisation rather than (λ, β). The log-transform symmetrises the Gamma, making the marginal of $\ln \lambda$ approximately Normal. In this space, the joint posterior is much closer to bivariate Normal, and the Hessian-derived covariance provides a better representation of the full posterior — including the tails. This is consistent with the standard practice of reporting the PMLM covariance in $(\ln \lambda, \beta)$ space (see Section 4), but the *reason* for this choice is rarely articulated: it is not merely a matter of convenience, but a consequence of the fact that the Gamma posterior for the rate becomes approximately Normal under the log-transform.

Important: When moving to high-order discretization ($n \geq 5$ points), the parametric form of the joint posterior matters. The eigendecomposition approach is only exact for multivariate Normal posteriors. For (λ, β), where the rate marginal is Gamma, the eigendecomposition should be performed in $(\ln \lambda, \beta)$ space to minimise distortion from the elliptical assumption. Even then, the approximation degrades in the extreme tails.

9.1.5 From Posterior to Logic Tree: An Open Problem

It is important to be honest about the limitations of current methods for constructing logic trees from posterior distributions. When the Full Bayesian Bayesian model provides the full joint posterior as MCMC samples, one might hope to simply extract a logic tree representation from this *true* posterior distribution. In practice, this is not straightforward.

The fundamental difficulty is that a logic tree requires a small number of discrete nodes with associated weights, while the posterior is a continuous bivariate distribution. Converting one to the other requires solving a *discretization* problem: choosing node locations and weights such that some property of the continuous distribution is preserved in the discrete approximation. Different choices of what to preserve (moments, quantiles, CDF shape, Wasserstein distance) lead to different discretizations, and there is no universally optimal solution.

Several approaches have been investigated:

- **Grid with Voronoi cells**: Define a regular grid and assign each MCMC sample to its nearest node, with weights proportional to the fraction of samples in each cell. This is simple but can produce poor moment preservation and depends on arbitrary grid placement.
- **Moment-matching**: Choose nodes and weights to match the first few moments of each marginal. This works well for means and variances but does not guarantee accurate tail quantiles.
- **Quantile-based discretization**: Place nodes at specified quantiles (e.g., P5, P50, P95). This preserves specified marginal quantiles but does not necessarily preserve the joint structure.
- **Optimal transport**: Minimise the Wasserstein distance between the discrete and continuous distributions. This has good theoretical properties but is computationally expensive and sensitive to the choice of metric.

None of these approaches is entirely satisfactory for the joint distribution. Each trades off accuracy in one aspect (moments, quantiles, tail behaviour, correlation) against accuracy in others. This remains an open area of research and practitioners should be aware that the choice of discretization method introduces its own epistemic uncertainty — separate from, and in addition to, the parametric uncertainty being discretized.

Posterior Shape and Discretization Recommendations:

1. When using PMLM or other classical methods that provide only point estimates and variances, assume a **Gamma distribution** for the activity rate λ and a **Normal distribution** for the slope β
2. For 3-point logic tree discretization, the choice of distributional form is not critical — the Normal assumption is adequate for both parameters
3. For 5-point or higher discretization, perform the eigendecomposition in $(\ln\lambda, \beta)$ space to exploit the approximate normality of $\ln\lambda$, and be aware that tail accuracy will depend on the assumed parametric form
4. The Full Bayesian provides the full joint posterior, but converting this to a logic tree remains non-trivial — no single discretization method is universally optimal, and the choice of method is itself a source of epistemic uncertainty

9.1.6 Eigendecomposition Approach for Correlated Parameters

When parameters are correlated, simply discretizing each marginal independently and taking all combinations produces incorrect joint weights. The proper approach uses eigendecomposition to identify independent components, discretizes these, then transforms back to the original parameter space.

Given correlated parameters (L, β) where $L = \ln(\lambda)$, with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$:

1. **Eigendecompose**: $\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top$ where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2)$ contains eigenvalues and $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2]$ contains eigenvectors
2. **Define independent variables**: $\boldsymbol{z} = \boldsymbol{V}^\top(\boldsymbol{x} - \boldsymbol{\mu})$ which have $\mathrm{Cov}(\boldsymbol{z}) = \boldsymbol{\Lambda}$ (diagonal)
3. **Discretize each z_i independently** using a standard n-point scheme (e.g., Miller & Rice (1983) nodes at $\pm\sqrt{3}\sigma_i$ with weights $[1/6, 2/3, 1/6]$, or the EPT nodes and weights from Keefer & Bodily (1983))
4. **Transform back**: $\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{V}\boldsymbol{z}$

This ensures marginal moments and correlation structure are preserved with mathematically justified weights, avoiding ad-hoc likelihood normalisation approaches.

9.1.6.1 Relationship to Rectangular Grids in Parameter Space

The eigendecomposition produces a grid that is rectangular in the principal-axis coordinates $\boldsymbol{z}$, but *rotated* relative to the original $(\ln\lambda, \beta)$ axes. This differs from current practice, where a rectangular grid is defined in the original parameter space and weights are assigned by normalising the bivariate density evaluated at the grid nodes. Practitioners accustomed to the latter approach may question whether the rotated grid is necessary.

The answer is straightforward: ad-hoc weight assignment on a rectangular grid in parameter space does not, in general, preserve the moments of the target distribution. In current

UK practice, the Hessian-derived covariance is used only to extract standard deviations (and implicitly correlations) for defining the grid node locations; the weights themselves are then computed by normalising the penalised log-likelihood evaluated at each grid node (see Equation 4.64). This is not the same as evaluating the posterior density, nor is it the same as computing probability masses that would reproduce the correct mean, variance, and covariance. Using penalised log-likelihood values as a basis for weights has no formal statistical justification — the penalised log-likelihood is not a probability density, and normalising it does not yield valid probability masses. If anything, this approach is likely to distort the correlation structure further than a density-based scheme would, because the penalty term introduces an additional asymmetry that has no relationship to the joint probability content of each grid cell.

There are three ways to handle correlation on a discrete grid:

1. **Likelihood-based weights on a rectangular grid**: This is the approach most commonly encountered in UK practice, where normalised penalised log-likelihood values are used as weights (see Equation 4.64). As noted in Section 4, there is no mathematical basis for this approach — it does not preserve moments and introduces uncontrolled distortion of the correlation structure.
2. **Moment-matching optimisation**: Given fixed node locations on a rectangular grid, one can solve for weights that exactly reproduce the target moments. This is feasible in principle, but the system is underdetermined (9 weights, fewer moment constraints) and solutions are not unique. For strong correlations or poorly placed nodes, some solutions may require negative weights, which are unphysical in a logic tree context.
3. **Conditional discretization**: Discretize one marginal first (e.g., $\ln \lambda$), then for each node discretize the conditional distribution of β given $\ln \lambda$. This produces a non-rectangular grid — the β nodes shift with each $\ln \lambda$ value — but preserves moments exactly and guarantees positive weights. For a bivariate Normal, this is:

$$\begin{aligned} L_i &= \mu_L + z_i \sigma_L \\ \beta_j \mid L_i &= \mu_\beta + \rho \frac{\sigma_\beta}{\sigma_L}(L_i - \mu_L) + z_j \sigma_\beta \sqrt{1-\rho^2} \end{aligned} \tag{9.1}$$

   This conditional approach is mathematically equivalent to the eigendecomposition for bivariate Normal distributions. It may be more intuitive to practitioners because it operates directly in the parameter space, although the resulting grid is a parallelogram rather than a rectangle.

In summary, there is no simple way to make a rectangular grid in the original parameter space correctly represent a correlated bivariate distribution — whether using likelihood-based weights (as in current practice) or density-based weights. The correlation must be

encoded either through grid rotation (eigendecomposition), through shifted node locations (conditional discretization), or through constrained weight optimisation — and of these, the first two are both exact and always produce positive weights.

Figure 9.3 illustrates these three approaches on a representative bivariate distribution with the parameters used throughout this section ($\sigma_{\ln\lambda} = 0.30$, $\sigma_b = 0.15$, $\rho = -0.5$). The conditional approach (panel a) aligns nodes with the density contours and preserves all moments exactly. The density-based approach (panel b) places nodes on a rectangular grid with weights proportional to the joint density — the resulting weights do not preserve the covariance. The moment-matching approach (panel c) solves for weights that reproduce target moments, but produces negative weights for the off-diagonal nodes, which are unphysical in a logic tree context. Panels (d–f) compare the weight distributions and moment preservation errors across the three methods.

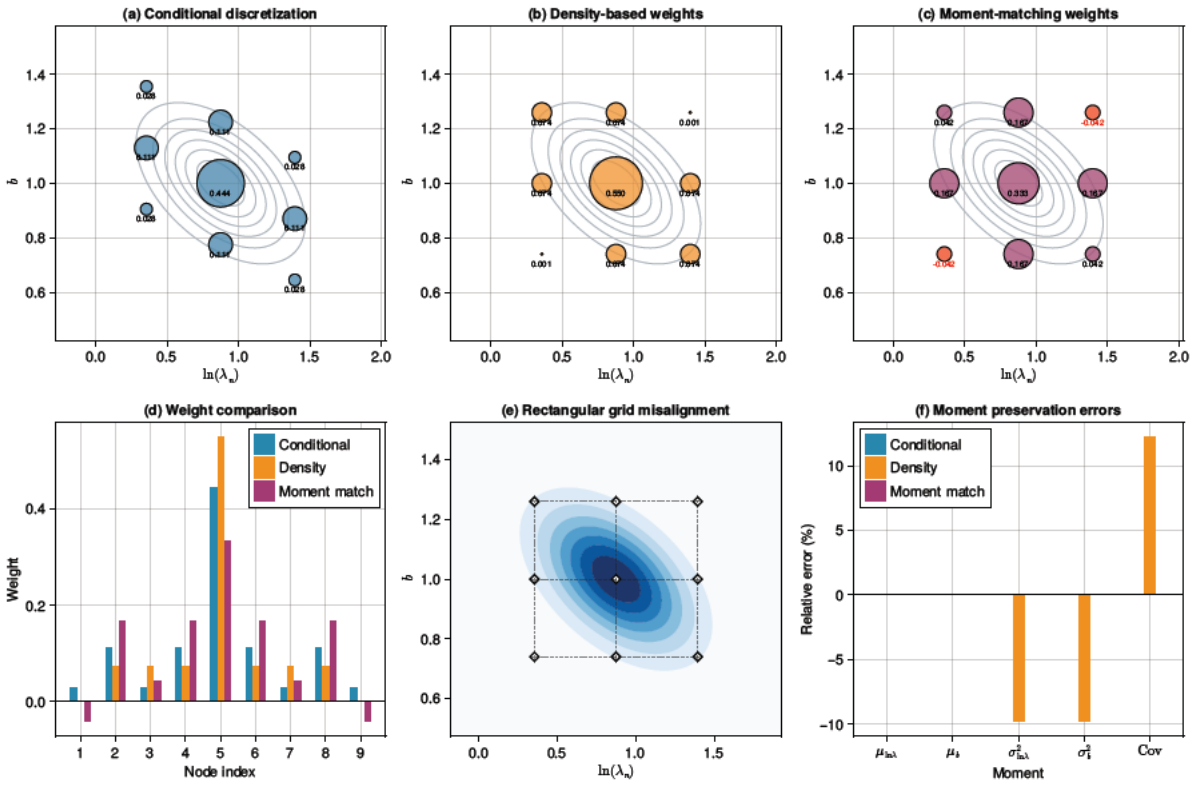

Figure 9.3: Comparison of discretization approaches for correlated parameters ($\sigma_{\ln\lambda} = 0.30$, $\sigma_b = 0.15$, $\rho = -0.5$). (a) Conditional discretization preserves moments by aligning nodes with correlation structure. (b) Density-based weights on a rectangular grid do not preserve moments. (c) Moment-matching finds optimal weights but can produce negative values (shown in red). (d) Weight distributions differ substantially across methods. (e) Rectangular grids are misaligned with density contours. (f) Moment preservation errors: conditional discretization is exact; density-based and moment-matching approaches introduce errors of differing character.

9.1.6.2 Handling the m_n to $m_{\min}$ Transformation

A critical question arises: should the eigendecomposition be performed at m_n (where the PMLM fit is obtained) or at $m_{\min}$ (where hazard calculations are performed)?

Two approaches are possible:

Approach 1 (Transform grid):

1. Eigendecompose $\boldsymbol{\Sigma}$ at m_n where correlation is positive
2. Create the rectangular grid aligned with principal axes at m_n
3. Transform each node to $m_{\min}$ using $\ln \lambda(m_{\min}) = \ln \lambda_n - \beta \Delta m$
4. Keep original weights

Approach 2 (Transform covariance first):

1. Transform the distribution to $m_{\min}$ using the linear transformation $\boldsymbol{T} = \begin{pmatrix} 1 & -\Delta m \\ 0 & 1 \end{pmatrix}$:

$$\boldsymbol{\Sigma}_{m_{\min}} = \boldsymbol{T} \boldsymbol{\Sigma}_{m_n} \boldsymbol{T}^{\top} \tag{9.2}$$

2. Eigendecompose $\boldsymbol{\Sigma}_{m_{\min}}$ where correlation is negative
3. Create rectangular grid aligned with principal axes at $m_{\min}$

Equivalence of Approaches:

Both approaches produce grids with identical effective moments (mean, variance, correlation) at $m_{\min}$. The transformation $W = L - \beta \Delta m$ is linear, so covariance structure is preserved regardless of when discretization occurs.

However, Approach 2 produces a rectangular grid aligned with the correlation structure at $m_{\min}$, which is conceptually cleaner and easier to interpret.

9.1.6.3 Practical Recommendation

For logic tree discretization at hazard reference magnitude $m_{\min}$:

1. **Obtain PMLM outputs** at m_n: $\left(\hat{\lambda}_n, \hat{\beta}, \boldsymbol{\Sigma}\right)$
2. **Transform** the distribution to $m_{\min}$:
   - Mean: $\mu_W = \ln\left(\hat{\lambda}_n\right) - \hat{\beta} \Delta m$, $\mu_\beta = \hat{\beta}$
   - Covariance: $\boldsymbol{\Sigma}_{m_{\min}} = \boldsymbol{T} \boldsymbol{\Sigma}_{m_n} \boldsymbol{T}^{\top}$ with $\boldsymbol{T} = \begin{pmatrix} 1 & -\Delta m \\ 0 & 1 \end{pmatrix}$
3. **Eigendecompose** at $m_{\min}$: $\boldsymbol{\Sigma}_{m_{\min}} = \boldsymbol{V} \boldsymbol{\Lambda} \boldsymbol{V}^{\top}$
4. **Discretize** along principal axes using Miller-Rice or the optimised schemes discussed below
5. **Transform back** to $(\ln \lambda, \beta)$ coordinates: $\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{V} \boldsymbol{z}$

This approach:

- Has rigorous mathematical foundation (preserves moments)
- Correctly handles the correlation sign change from positive at m_n to negative at $m_{\min}$
- Produces a rectangular grid aligned with uncertainty structure at $m_{\min}$
- Avoids ad-hoc likelihood normalisation

Technical Detail: Current UK Practice Is Likely Not Catastrophically Wrong

Current UK practice computes weights by normalising the penalised log-likelihood evaluated at grid nodes defined at m_n. As discussed above, this has no formal statistical justification and does not correctly preserve the moments of the target distribution. Nevertheless, the approach is unlikely to be catastrophically wrong, for the following reasons:

1. The likelihood-based weights at m_n attempt to capture the positive correlation there, and while they do not reproduce it exactly, the qualitative pattern (higher weight for combinations where λ_n and β move together) is at least directionally correct
2. When nodes are transformed to $m_{\min}$ (each (λ_n, β) becoming $(\lambda_n e^{-\beta\Delta m}, \beta)$), the differential transformation of λ nodes based on β implicitly encodes the correlation sign change — this step is internally consistent regardless of whether the starting weights are exact
3. The resulting effective moments at $m_{\min}$ will approximately match the analytically transformed distribution, though with some distortion inherited from the inexact weight assignment at m_n

The overall assessment is that the current approach is wrong but probably not by a large amount for typical UK parameter ranges. However, this is not a compelling reason to continue using it. The eigendecomposition and conditional discretization approaches described above are no more difficult to implement, are mathematically exact, and do not require practitioners to rely on precedent as a substitute for justification. The fact that an unjustified method happens to produce roughly acceptable results does not confer any advantage over a justified method that produces exact results.

9.2 Optimal Discretization for Logic Trees

Standard discretization schemes may not be optimal for fractile estimation. In UK regulatory practice, design decisions are often influenced by fractile hazard — particularly the 84th percentile ground motion — which depends on the upper tail of the distribution of seismicity parameters propagated through the logic tree. If the discrete representation of parametric uncertainty does not faithfully reproduce the tail of the rate distribution, the resulting fractile hazard estimates will be biased regardless of how well the mean hazard is preserved.

To investigate this, the accuracy of different discretization schemes was assessed using a controlled numerical experiment. The setup is as follows:

- **Parameters:** A bivariate Normal distribution over $(\ln\lambda, \beta)$ with $\sigma_{\ln\lambda} = 0.30$, $\sigma_\beta = 0.15$, and correlation $\rho = -0.5$, representative of typical UK source zone posteriors at the hazard reference magnitude $m_{\min}$.
- **Quantity evaluated:** The Gutenberg-Richter exceedance rate $\lambda(M) = \exp(\ln\lambda - \beta(M - M_{\min}))$ at magnitudes $M = 4.0$ to 6.5 in steps of 0.5 units, with $M_{\min} = 4.0$.

- **Reference:** Monte Carlo simulation with 100,000 samples from the bivariate distribution, providing reference values for the mean rate and 84th percentile rate (P84) at each magnitude.
- **Error metric:** Percentage error of the discrete approximation relative to the Monte Carlo reference, i.e., $100 \times \frac{\text{discrete} - \text{reference}}{\text{reference}}$, evaluated for both the mean rate and P84 rate.
- **Correlation handling:** All discretization schemes use conditional discretization to preserve the negative correlation structure, as described in Section 9.1.6.

It is important to note that this experiment evaluates the discretization of seismicity parameters only — it does not involve ground-motion models or full hazard calculations. The errors reported below relate to how well a discrete set of (λ, β) branches reproduces the distribution of exceedance rates implied by the continuous bivariate posterior. In a full PSHA, additional sources of discretization error from other logic tree nodes (e.g., ground-motion model branches, maximum magnitude) would compound these effects. The results therefore represent a lower bound on the total discretization error in the seismicity component of the logic tree.

9.2.1 The Mean-Fractile Trade-off

Standard discretization schemes were derived from different theoretical perspectives:

Scheme	z-values	Weights	Basis
Miller-Rice	± 1.732	[0.167, 0.667, 0.167]	Gaussian quadrature
Ext. Pearson-Tukey	± 1.645	[0.185, 0.630, 0.185]	P5/P50/P95 percentiles
Ext. Swanson-Megill	± 1.282	[0.300, 0.400, 0.300]	P10/P50/P90 percentiles

Miller-Rice is mathematically optimal for expected value calculations, but none of these schemes were designed to optimise fractile estimation accuracy.

Testing revealed a fundamental finding: all standard schemes achieve similar P84 accuracy (~6–8% error) despite their different theoretical bases (Figure 9.4). Miller-Rice achieves excellent mean accuracy (0.15% error) but poor P84 accuracy (7.6% error). The percentile-based methods offer marginal improvements in P84 but sacrifice mean accuracy — the trade-off is systematic and unavoidable for 3-point discretizations.

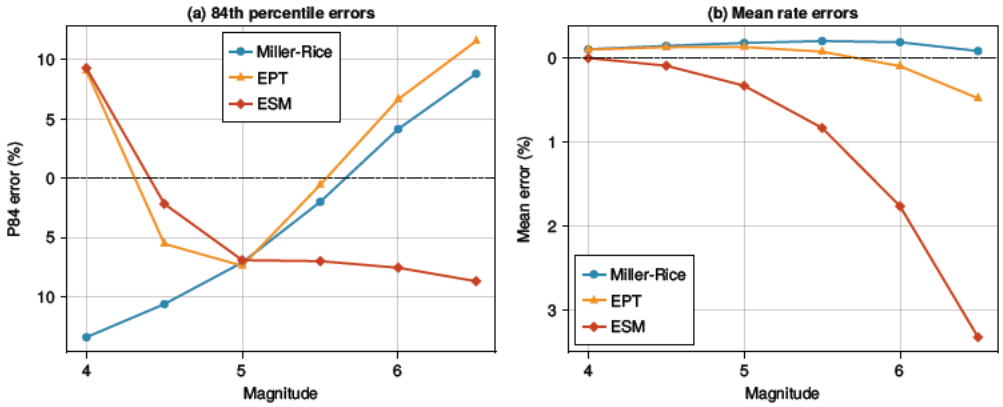

Figure 9.4: Comparison of standard 3-point discretization schemes for exceedance rate estimation. (a) 84th percentile errors by magnitude. (b) Mean rate errors by magnitude. All three schemes produce similar P84 errors ($\sim$ 6–8%) despite different theoretical bases, while mean errors are generally smaller and differ more across schemes.

This occurs because the schemes share a common structural feature: heavy weight on the median branch ($w_{\text{mid}} = 0.63$–0.67), which creates large gaps in the cumulative weight distribution near the 84th percentile. Sweeping over the full space of variance-preserving 3-point schemes reveals a continuous trade-off between mean and fractile accuracy (Figure 9.5): schemes optimised for mean accuracy (Miller-Rice) perform poorly for P84, while schemes that improve P84 performance necessarily sacrifice mean accuracy.

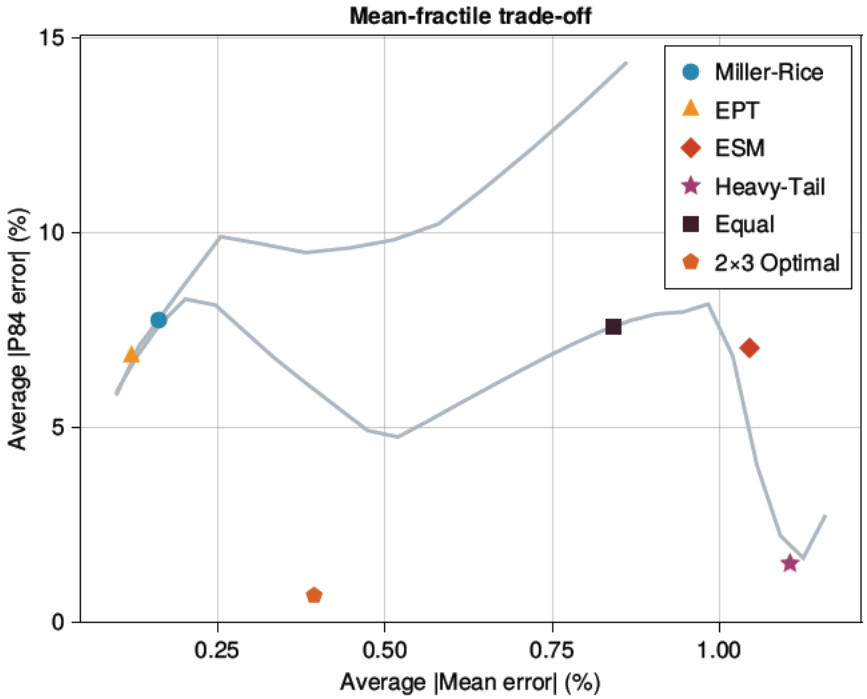

Figure 9.5: The mean-fractile trade-off in 3-point discretization. The grey curve traces all variance-preserving schemes (parameterised by z); named schemes are marked. Miller-Rice sits at the low-mean-error end with high P84 error. The Heavy-Tail and 2×3 Optimal schemes (introduced in subsequent sections) break out of the 3×3 trade-off curve by redistributing weight or using asymmetric grids.

Technical Detail: Why Standard Schemes Fail for P84

Fractiles are computed by inverting the CDF. The interpolation error depends on the gap in cumulative weights around 0.84. For Miller-Rice (3×3 grid), cumulative weights near P84 are:

$$[\ldots, 0.583, 0.694, 0.833, 1.000] \tag{9.3}$$

P84 falls in the [0.833, 1.000] gap of width 0.167, requiring interpolation 4% into a 17% span. The large gap arises because the heavy median weight (0.667) concentrates probability mass away from the tails.

9.2.2 Numerical Optimisation

Direct numerical optimisation to minimise P84 error while preserving variance yields a radically different scheme:

Scheme	z	Weights	P84 Error
Miller-Rice	1.732	[0.167, 0.667, 0.167]	7.6%
Ext. Pearson-Tukey	1.645	[0.185, 0.630, 0.185]	6.7%
Heavy-Tail (optimal)	**1.034**	**[0.468, 0.064, 0.468]**	**1.5%**

The Heavy-Tail scheme achieves a 5-fold reduction in P84 error by placing only 6% weight on the median branch. Figure 9.6 compares the standard and optimised schemes: panels (a) and (b) show the P84 and mean errors by magnitude, while panel (c) reveals the mechanism — the Heavy-Tail scheme creates a cumulative weight structure with much smaller gaps around the 84th percentile, enabling more accurate interpolation of the P84 rate.

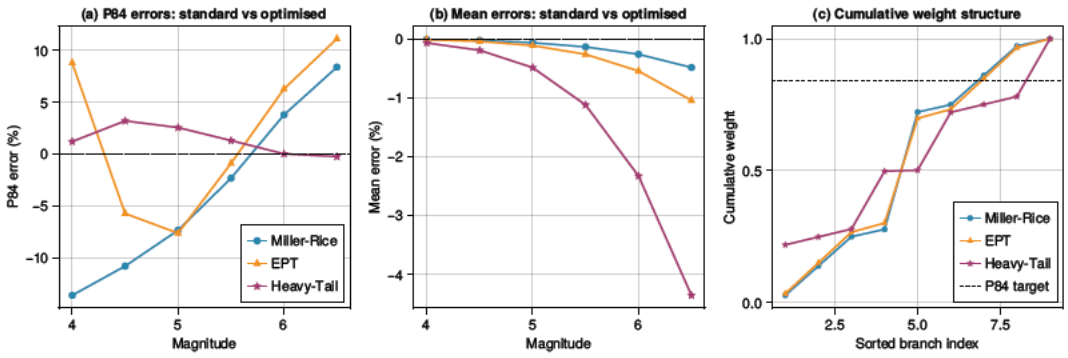

Figure 9.6: Performance of the optimised Heavy-Tail discretization compared to standard schemes. (a) P84 errors by magnitude: Heavy-Tail reduces error from $\sim 7\%$ to $\sim 1.5\%$. (b) Mean errors: the Heavy-Tail scheme sacrifices some mean accuracy ($\sim 1\%$) relative to Miller-Rice ($\sim 0.15\%$). (c) Cumulative weight structure at $M = 5.5$: Heavy-Tail places cumulative weight steps more uniformly, avoiding the large gap near $P = 0.84$ that plagues standard schemes.

9.2.3 Asymmetric Discretization: The 2×3 Grid

For Gutenberg-Richter parameters, the b-value sensitivity is 5–6 times larger than the $\ln \lambda$ sensitivity at $M \approx 6$. This suggests allocating more points to the more influential parameter.

The optimised 2×3 scheme (2 points for $\ln \lambda$, 3 heavy-tail points for b-value) achieves:

Structure	Branches	|Mean|	|P84|	Cost
3×3 Miller-Rice	9	0.15%	7.6%	100%
3×3 Heavy-Tail	9	1.17%	1.5%	100%
2×3 Optimal	**6**	**0.18%**	**0.50%**	**67%**

The optimised 2×3 scheme achieves **three times better P84 accuracy than the 3×3 Heavy-Tail scheme** while using **33% fewer branches**. Figure 9.7 shows the full comparison: the 2×3 grid matches or exceeds the 3×3 Miller-Rice scheme for mean accuracy while dramatically outperforming it for P84, and the summary panel (c) confirms that no 3×3 configuration achieves the combined cost-accuracy performance of the asymmetric grid.

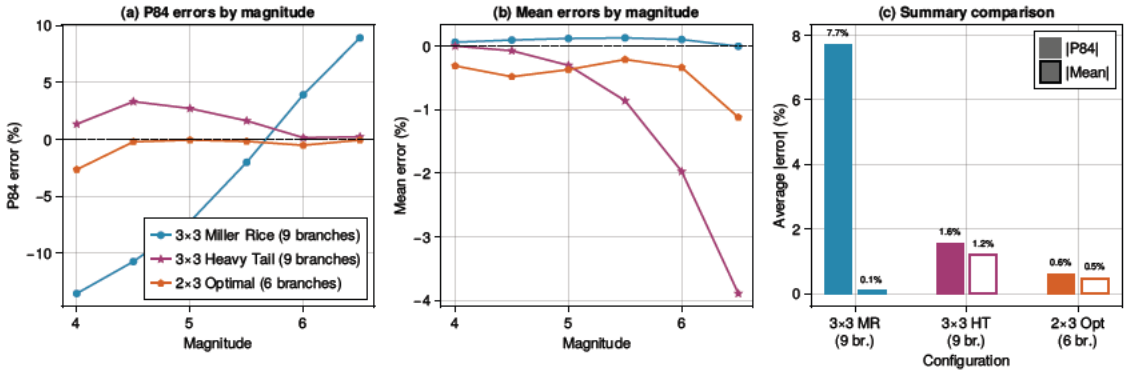

Figure 9.7: Performance of the asymmetric 2×3 discretization (2 points for $\ln \lambda$, 3 heavy-tail points for b-value) compared to symmetric 3×3 grids. (a) P84 errors by magnitude. (b) Mean errors by magnitude. (c) Summary comparison of average absolute errors: the 2×3 scheme achieves the best combined accuracy with only 6 branches (vs. 9 for the 3×3 grids).

9.2.4 Weight Distributions in Large Logic Trees

The preceding sections analysed discretization for a single pair of correlated parameters. In practice, seismic hazard logic trees combine multiple nodes – source geometry alternatives, seismicity parameters for each zone, ground-motion model branches, maximum magnitude assessments – and the total number of end branches is the product across all nodes. For a tree with n nodes each having k branches, there are k^n end branches, each with weight equal to the product of marginal weights along its path:

$$w_{\text{branch}} = \prod_{i=1}^{n} w_i \tag{9.4}$$

Taking logarithms converts this product to a sum: $\ln w_{\text{branch}} = \sum_{i=1}^{n} \ln w_i$. By the Central Limit Theorem, as n increases the distribution of log-weights converges to a Normal distribution, so that branch weights follow an approximately log-normal distribution. This has important practical consequences for the behaviour of large logic trees.

Figure 9.8 illustrates three effects. First, weight dispersion grows exponentially with tree size (panel a): the ratio of maximum to minimum branch weight depends on the max/min ratio of the marginal weights at each node, and this ratio compounds multiplicatively across nodes. Any scheme with unequal marginal weights — whether 3-point Miller-Rice ($w_{\max}/w_{\min} = 4$), the Heavy-Tail scheme ($w_{\max}/w_{\min} \approx 7$), or higher-order quadrature rules — will produce exponentially growing dispersion, with the rate determined by the marginal weight ratio. By contrast, equal marginal weights produce identical branch weights regardless of tree size. Second, for moderate tree sizes ($n = 4$), the distribution of log-weights is already approximately bell-shaped for non-equal schemes, but with very different spreads (panel b). Third, for $n = 6$ the Heavy-Tail log-weight distribution is well-described by the theoretical Normal predicted by the CLT (panel c), confirming the log-normal convergence.

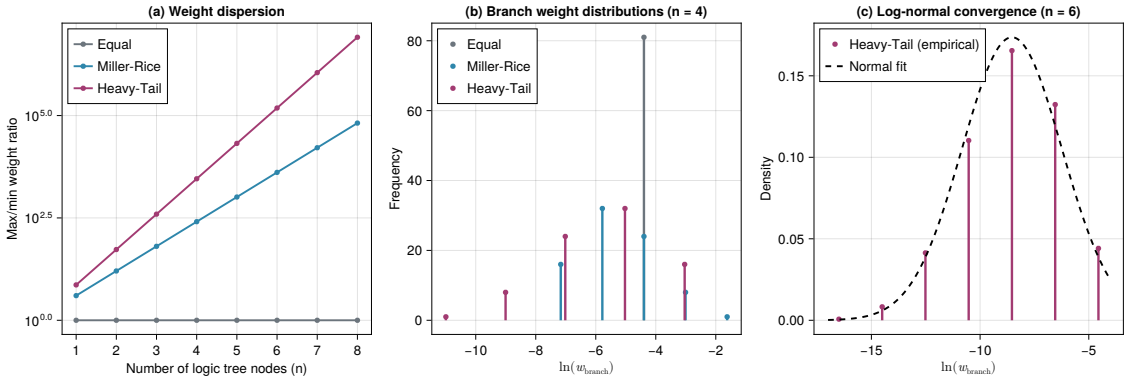

Figure 9.8: Behaviour of weight distributions in large logic trees. (a) Max/min branch weight ratio grows exponentially with tree size for any scheme with unequal marginal weights; the rate of growth depends on the marginal weight ratio at each node. Equal weights maintain a ratio of 1:1 regardless of tree depth. (b) Distribution of log branch weights for $n = 4$ nodes, showing the different spreads produced by Equal, Miller-Rice, and Heavy-Tail marginal weights. (c) For $n = 6$ nodes with Heavy-Tail weights, the empirical log-weight distribution closely matches the Normal distribution predicted by the Central Limit Theorem (dashed curve).

The extreme weight dispersion from non-equal schemes has two practical implications. First, branches with negligible weight contribute almost nothing to the hazard calculation but still incur full computational cost — in a tree with 10^5-fold weight dispersion, most of the computational budget is wasted on branches that are effectively irrelevant. Second, the fractile accuracy of the discrete approximation depends on how cumulative weights align with the target percentile, and extreme dispersion can create irregular gaps in the cumulative weight function near the target. Practices for fractile estimation vary across PSHA codes — some interpolate between bounding cumulative weights, others take the first branch whose cumulative weight meets or exceeds the target — and extreme weight dispersion creates problems for both approaches. The interaction between weight structure, branch ordering, and fractile estimation method warrants further investigation, but is beyond the scope of this report.

9.2.5 Discretization Recommendations

On the basis of the results presented above, the following observations are made:

- Standard discretization schemes (Miller-Rice, EPT, ESM) were designed to preserve moments or match specific quantiles of the marginal distribution. None were designed with the goal of enabling stable estimation of hazard fractiles, and all produce similar P84 errors of 6–8% for the seismicity parameters examined here (Figure 9.4).
- The Heavy-Tail scheme ($z = \pm 1.03$, $w = [0.47, 0.06, 0.47]$) reduces P84 error to 1.5% while maintaining acceptable mean accuracy (~1%) (Figure 9.6).
- For Gutenberg-Richter parameters specifically, the asymmetric 2×3 grid achieves 0.5% P84 error with only 6 branches — the best cost-accuracy trade-off identified (Figure 9.7).
- The choice of discretization scheme is a source of epistemic uncertainty, and the numerical results for hazard fractiles have a non-trivial dependence on the discretization scheme. Sensitivity analyses should compare fractile estimates from different approaches (Figure 9.5).

10 Theme B: Regional Analysis and Zone Correlations

UK seismic hazard studies partition the region into area source zones whose boundaries are based primarily on geological regime, geophysical character, and precedent, with observed seismicity patterns playing a supporting role. The 2020 UK National Seismic Hazard Model (Mosca et al., 2022) uses 22 source zones that vary greatly in size, data content, and geological setting — from the well-instrumented onshore zones with hundreds of recorded events to offshore zones with fewer than ten. Despite this heterogeneity, current practice applies the same PMLM prior on b-value to every zone: the Johnston et al. (1994) global SCR prior with $b = 1.0$ and $\sigma_b \approx 0.087$.

As established in Section 6, this PMLM procedure is mathematically equivalent to MAP estimation with a Normal prior on β, where the prior standard deviation is determined by the penalty weight. The practical consequence is that for data-sparse zones — which constitute the majority of UK sources — the prior dominates the posterior estimate, pulling zone-specific b-values toward the global mean of 1.0. This creates a homogenising effect: zones that may genuinely differ in their seismicity characteristics are assigned near-identical parameter estimates, not because the data support this conclusion, but because the prior overwhelms limited observations.

This situation raises three inter-related questions that motivate the investigations in this section:

1. *Is the PMLM prior reasonable for the UK as a whole?* The Johnston prior derives from a global compilation of stable continental region earthquakes. Whether $b = 1.0$ is appropriate for the UK specifically — and whether UK seismicity is sufficiently homogeneous to be characterised by a single b-value — requires empirical assessment against the national catalogue.
2. *Is the same prior appropriate across all individual zones?* Zones span very different tectonic settings and data volumes. A prior that is informative but appropriate for a well-recorded onshore zone may be severely misspecified for a data-sparse offshore zone, or for zones whose frequency-magnitude distributions are complicated by bipartite behaviour or completeness artifacts.
3. *With a prior this strong, is the 22-zone subdivision justified?* If most zone estimates are dominated by the prior rather than by local data, the effective number of independent parameters is far fewer than the nominal count. The fine subdivision may create an illusion of spatial resolution that the data cannot support.

These questions have direct consequences for logic tree construction. With s source zones each discretised into a $k \times k$ grid of GR parameters, the full logic tree formally contains $(k^2)^s$ branches. Current practice treats zone parameters as independent across branches, but this is inconsistent with the estimation procedure: the shared prior that constrains zone i's b-value equally constrains zone j's, creating implicit positive correlations between zone estimates. Treating correlated parameters as independent overstates the true epistemic uncertainty in the aggregated hazard — compounding the discretization issues identified in Section 9.2. The correlation structure is further complicated by physical constraints on total moment release and by uncertainty in zone boundaries, both of which introduce additional inter-zone dependencies.

The remainder of this section addresses these issues in turn. First, we estimate a regional b-value from the pooled UK catalogue to provide a data-driven benchmark against which the PMLM prior can be assessed. We then test whether individual zone estimates are statistically distinguishable from this regional value and from each other, before examining the correlation structure that arises from shared priors, moment constraints, and boundary uncertainty. Finally, we present a systematic diagnostic assessment of PMLM prior appropriateness for each UK source zone.

10.1 Regional b-value from Pooled UK Catalogue

A fundamental question is whether individual source zones genuinely exhibit different seismicity parameters, or whether observed variation is simply sampling variability from a common distribution. To investigate this, the complete UK earthquake catalogue was analysed to obtain a regional b-value estimate that serves as a data-driven benchmark against which individual zone estimates and the PMLM prior can be assessed.

The BGS earthquake catalogue was processed following standard practice: completeness filtering using the model of Musson (2007), declustering using the algorithm of Burkhard & Grünthal (2009), and magnitude conversion from M_L to M_w using the quadratic relationship of Grünthal et al. (2009) (Equation 7.2). Two estimation methods were applied to the pooled catalogue:

1. **Weichert MLE**: The maximum likelihood procedure of Weichert (1980), which properly accounts for variable observation periods across magnitude bins but does not account for magnitude measurement uncertainties. This is equivalent to the PMLM procedure with zero penalty ($W = 0$) and therefore provides an unpenalised reference estimate.
2. **Full Bayesian L5**: The full Bayesian model developed in Section 6, which accounts for magnitude measurement uncertainty (per-event σ_{M_L}), magnitude conversion error, and

selection effects via a marginalised likelihood implemented in Stan. This is the analysis approach that this report advocates as most rigorous.

Results are presented for two minimum magnitude thresholds (Table 10.1): $M_w \geq 3.0$, the lowest completeness level in the Musson (2007) model (complete from 1975), and $M_w \geq 2.5$, which exploits the extended instrumental completeness identified by Mosca et al. (2022). Both thresholds are defined directly in the M_w scale.

Table 10.1: Regional Gutenberg-Richter parameter estimates from the pooled UK catalogue with Grünthal et al. (2009) magnitude conversion. The Weichert MLE ignores magnitude uncertainty; the Full Bayesian L5 accounts for per-event measurement uncertainty, conversion error, and selection effects. $\lambda(3.0)$ is the annual rate above $M_w = 3.0$; all rates are reported at this common reference magnitude regardless of the fitting threshold $M_{w,\min}$, to enable direct comparison. The $M_{w,\min} = 2.5$ rows use additional lower-magnitude data for parameter estimation but extrapolate the fitted rate to the same reference level. N is the number of mainshock events after completeness filtering and declustering.

$M_{w,\min}$	Method	N	b	$\lambda(3.0)$	$\rho(\lambda, \beta)$
2.5	Weichert MLE	373	0.96 ± 0.03	2.32 ± 0.14	0.52
2.5	Full Bayesian	373	0.86 ± 0.03	1.59 ± 0.09	-0.46
3.0	Weichert MLE	206	0.93 ± 0.05	2.18 ± 0.18	0.56
3.0	Full Bayesian	206	0.85 ± 0.04	1.45 ± 0.10	0.06

The two methods yield strikingly different b-value estimates. The Weichert MLE gives $b = 0.96 \pm 0.03$ ($M_w \geq 2.5$) and $b = 0.93 \pm 0.05$ ($M_w \geq 3.0$) — both within approximately 1σ of the Johnston PMLM prior mean of $b = 1.0$. However, the Full Bayesian L5 model gives substantially lower estimates: $b = 0.86 \pm 0.03$ ($M_w \geq 2.5$) and $b = 0.85 \pm 0.04$ ($M_w \geq 3.0$). The difference of approximately 0.08–0.10 in b is a direct consequence of the Eddington bias correction: methods that ignore magnitude uncertainty overestimate b because events with true magnitudes below the completeness threshold can scatter upward into the analysis, preferentially populating the lower magnitude bins and steepening the apparent frequency-magnitude slope. The Full Bayesian model corrects for this by marginalising over the unknown true magnitudes. The lower rate estimates from the Full Bayesian ($\lambda(3.0) \approx 1.5$ vs 2.2–2.3 events/yr) similarly reflect the removal of scatter-IN events that inflate the apparent rate. The modest differences in $\lambda(3.0)$ between the two fitting thresholds (e.g. 2.18 vs 2.32 for Weichert) arise because all rates are reported at the common reference magnitude $M_w = 3.0$: the $M_{w,\min} = 2.5$ analysis uses more data for parameter estimation, producing slightly different fitted parameters that yield a slightly different extrapolated rate at the same reference level.

This method sensitivity has important implications for the assessment of the PMLM prior. If the Weichert MLE is taken as the reference — as is standard practice in UK hazard studies — the prior mean of $b = 1.0$ appears reasonable. If instead the Full Bayesian is adopted, the prior mean sits approximately 1.7σ above the data-driven estimate, indicating moderate

tension. The choice of analysis method therefore determines whether the PMLM prior is viewed as appropriate or as mildly biased high.

10.2 Sensitivity to Magnitude Conversion

The choice of magnitude conversion has a significant effect on the estimated b-value (Table 10.2). This arises because the conversion relationships are nonlinear, causing the M_w scale to compress relative to M_L at lower magnitudes and expand at higher magnitudes. Since the b-value describes the rate of decay of event frequency with magnitude, any nonlinear transformation of the magnitude scale alters the apparent slope.

Table 10.2: Sensitivity of the regional b-value estimate to the choice of magnitude conversion, using a common threshold of $M \geq 3.0$ in the respective scale. The different event counts arise because the same M_L events map to different M_w values under each conversion, so different numbers exceed the $M_w 3.0$ threshold. All estimates use the Weichert MLE procedure. The minor differences in N and b for the Grünthal row relative to Table 10.1 reflect a slightly different completeness year (1970 vs 1975) used in the conversion sensitivity analysis.

Scale	b-value	σ_b	N
$M_L \geq 3.0$ (no conversion)	0.90	0.04	283
$M_w \geq 3.0$ (Stromeyer et al. 2004)	1.07	0.06	188
$M_w \geq 3.0$ (Grünthal et al. 2009)	0.97	0.06	195

The Grünthal relationship is preferred here as it represents an updated calibration over the Stromeyer et al. (2004) relationship. The spread of Weichert b-value estimates across the three scales ($b \approx 0.90$–1.07) illustrates that the choice of magnitude conversion introduces an epistemic uncertainty comparable in magnitude to the statistical uncertainty on b. With the preferred Grünthal conversion, the Weichert estimate of $b = 0.93 \pm 0.05$ is within 1σ of the Johnston PMLM prior mean of $b = 1.0$. However, as Table 10.1 shows, the Full Bayesian estimate ($b = 0.85$) is substantially lower, and the combined effect of magnitude conversion and analysis method spans a range of at least $b = 0.85$–1.07 — far wider than the statistical uncertainty from any single analysis. In practice, the conversion model would be treated as a separate epistemic branch in the logic tree, with each branch adopting its own parameter estimates, rather than attempting to absorb the conversion uncertainty into a wider prior on b.

10.2.1 Implications of the prior standard deviation

The PMLM prior standard deviation of $\sigma_b \approx 0.087$ adopted in UK practice merits further comment. The regional results in Table 10.1 show that the statistical uncertainty on the national b-value depends on both the minimum magnitude threshold and the analysis method: $\sigma_b = 0.03$ at $M_w \geq 2.5$ versus $\sigma_b = 0.04$–0.05 at $M_w \geq 3.0$. That the prior is increasingly informative as the threshold rises and data become sparser is not in itself problematic — this is precisely

how Bayesian updating should behave, with the prior exerting greater influence when data are limited.

The more fundamental question is whether a prior standard deviation that appears reasonable for the UK as a whole is also appropriate as a zone-level prior. The national σ_b of 0.03–0.06 reflects sampling uncertainty on a single regional b-value. A zone-level prior, by contrast, needs to capture the plausible range of b-values *across* individual zones — a fundamentally different quantity. If all UK zones genuinely share the same b-value (as the zone distinguishability tests below suggest for most onshore zones), then the zone-level prior should be narrow, reflecting minimal true variation. If zones exhibit genuine physical heterogeneity, the prior should be wider. The PMLM approach cannot distinguish between these scenarios: it applies the same $\sigma_b = 0.087$ regardless, which may be too wide if zones are truly homogeneous (over-stating uncertainty) or too narrow if some zones genuinely differ (under-stating it for those zones). A hierarchical model would learn the appropriate degree of zone-to-zone variability from the data, but the fixed PMLM prior cannot adapt in this way.

10.3 Zone Distinguishability Testing

The UK NSHM 2020 source model divides the study region into 22 seismic source zones, each assigned its own Gutenberg-Richter parameters. A natural question is whether the available earthquake data actually support this level of spatial resolution: can we distinguish the seismicity characteristics of adjacent zones, or does the zonation represent an over-parameterisation given the data? If most zones are statistically indistinguishable, a simpler source model with fewer zones – or a hierarchical model that partially pools parameters across zones – may be equally valid and more robust.

This question is particularly pressing for the UK, where earthquake catalogues are sparse by global standards. Many individual zones contain fewer than 20 events above the completeness threshold, making zone-specific parameter estimates highly uncertain. If the apparent zone-to-zone variation in b-value is largely sampling noise rather than genuine spatial heterogeneity, then treating each zone as independent wastes information and inflates parameter uncertainty unnecessarily.

Two complementary tests are applied. The first uses likelihood ratio tests to compare adjacent zone pairs directly, asking whether their b-values are significantly different from one another. The second compares each zone individually against the regional (pooled UK) b-value estimate, asking whether any zone deviates significantly from the national average.

10.3.1 Methodology

10.3.1.1 Data preparation

The analysis uses the BGS earthquake catalogue with events converted to moment magnitude using the Grünthal et al. (2009) relationship and filtered for completeness using the Musson (2007) model, as described in Section 5. The primary analysis uses a minimum magnitude of $M_w = 3.0$ (the lowest threshold in the standard completeness model, complete from 1970) and maximum magnitude of $M_w = 6.5$. A sensitivity analysis is also performed at $M_w = 2.5$ (also treated as complete from 1970) to assess whether the conclusions are robust to the inclusion of additional smaller events. Each event is assigned to the source zone containing its epicentre. Events falling outside all zone boundaries are excluded from the zone-specific analysis but are included in the pooled regional estimate.

10.3.1.2 Zone-specific parameter estimation

For each of the 22 source zones, the Gutenberg-Richter parameters (λ_n, β) are estimated by maximum likelihood using a binned Poisson model. Magnitudes are discretised into bins of width $\Delta m = 0.1$ over the range $[3.0, 6.5]$. The expected count in each bin is

$$\mu_i = \lambda_n \cdot \Delta F_i \cdot t_i \tag{10.1}$$

where ΔF_i is the probability mass assigned to bin i by the truncated exponential distribution and t_i is the observation period for that magnitude level (determined by the completeness model). The log-likelihood is

$$\ln L = \sum_i [n_i \ln \mu_i - \mu_i] \tag{10.2}$$

where n_i is the observed count in bin i. This formulation omits the factorial terms $\ln(n_i!)$, which are constant with respect to the parameters and cancel in likelihood ratio comparisons. Parameters are estimated by bounded optimisation (L-BFGS-B with $\lambda_n \in [0.001, 1000]$ and $\beta \in [0.5, 10]$), and uncertainties are obtained from the observed information matrix (inverse Hessian at the MLE).

10.3.1.3 Adjacency determination

Two zones are classified as adjacent if they share at least two vertices within a tolerance of $0.05°$ (approximately 4–5 km at UK latitudes). This geometric criterion identifies zone pairs that share a common boundary segment, as opposed to zones that merely touch at a single corner point. The tolerance accommodates minor coordinate rounding in the zone definitions.

10.3.1.4 Likelihood ratio test for shared b-value

For each adjacent zone pair, a likelihood ratio test (LRT) is used to test the null hypothesis that the two zones share a common b-value. Under the null hypothesis H_0, a single β is estimated jointly from both zones' data while allowing separate activity rates $\lambda_{n,1}$ and $\lambda_{n,2}$

(3 free parameters: $\lambda_{n,1}$, $\lambda_{n,2}$, β). Under the alternative H_1, each zone has its own β and λ_n (4 free parameters). The test statistic is

$$\Lambda = -2\left(\ln L_{H_0} - \ln L_{H_1}\right) \tag{10.3}$$

which, under standard regularity conditions, follows a χ^2 distribution with 1 degree of freedom (the single constraint $\beta_1 = \beta_2$). The null hypothesis is rejected at the 5% significance level if $p < 0.05$.

This test addresses only whether b-values differ. It is the relevant test for deciding whether it is reasonable to pool b-values across zones in a hierarchical model, but it does not answer the broader question of whether two zones can be merged into a single homogeneous source.

10.3.1.5 Full aggregation test

A more complete test asks whether two adjacent zones are genuinely distinguishable as seismic sources — that is, whether they differ in their area-normalised activity rate, their b-value, or both. Two zones with the same b-value but genuinely different seismicity densities (events per unit area per year) still require separate parameterisation in the source model; merging them would misrepresent the spatial distribution of seismicity even if the magnitude-frequency slope is identical.

Under the null hypothesis H_0 (zones are mergeable), both zones are treated as draws from a single homogeneous Poisson process with a common β and a common area-normalised rate. The zone-specific rates are constrained to be proportional to their areas:

$$\lambda_{n,i} = \frac{A_i}{A_{\text{total}}} \cdot \lambda_n^{\text{total}} \tag{10.4}$$

where A_i is the zone area (in km²) and $A_{\text{total}} = A_1 + A_2$. This gives 2 free parameters (λ_n^{total}, β). The alternative H_1 allows each zone its own λ_n and β (4 free parameters), giving $\mathrm{df} = 4 - 2 = 2$. Zone areas are computed from the polygon coordinates using the shoelace formula with latitude-dependent scaling (1° lon $\approx 111 \cos\varphi$ km, 1° lat ≈ 111 km).

10.3.1.6 Comparison to regional b-value

As a third test, each zone's MLE b-value is compared to the Weichert MLE regional estimate using a z-test:

$$z = \frac{b_{\text{zone}} - b_{\text{regional}}}{\sigma_{b,\text{zone}}} \tag{10.5}$$

where $\sigma_{b,\text{zone}}$ is the standard error of the zone's MLE b-value estimate. A two-sided p-value is computed from the standard normal distribution. This test identifies individual zones whose b-values are inconsistent with the national average, regardless of whether they happen to be adjacent to another zone. The reference value $b_{\text{regional}} = 0.97$ is the Weichert MLE from

the same zone aggregation pipeline used to estimate zone-specific parameters ($M_w \geq 3.0$, Grünthal conversion, completeness from 1970; see Table 10.2). This is slightly higher than the $b = 0.93$ in Table 10.1, which uses completeness from 1975, but ensures a like-for-like comparison: zone MLE estimates are tested against a regional MLE estimate from the same processing chain. We note that if the Full Bayesian estimate ($b = 0.85$, Table 10.1) were used as the reference instead, more zones would appear to deviate from the national average, since the reference itself would be substantially lower.

Figure 10.1 presents an overview of the zone-specific MLE b-value estimates with their confidence intervals, compared to the pooled regional estimate. The characteristic funnel shape — wider confidence intervals for zones with fewer events — is clearly evident in Figure 10.2, confirming that the apparent heterogeneity in zone b-values is predominantly a consequence of sampling variability rather than genuine spatial variation.

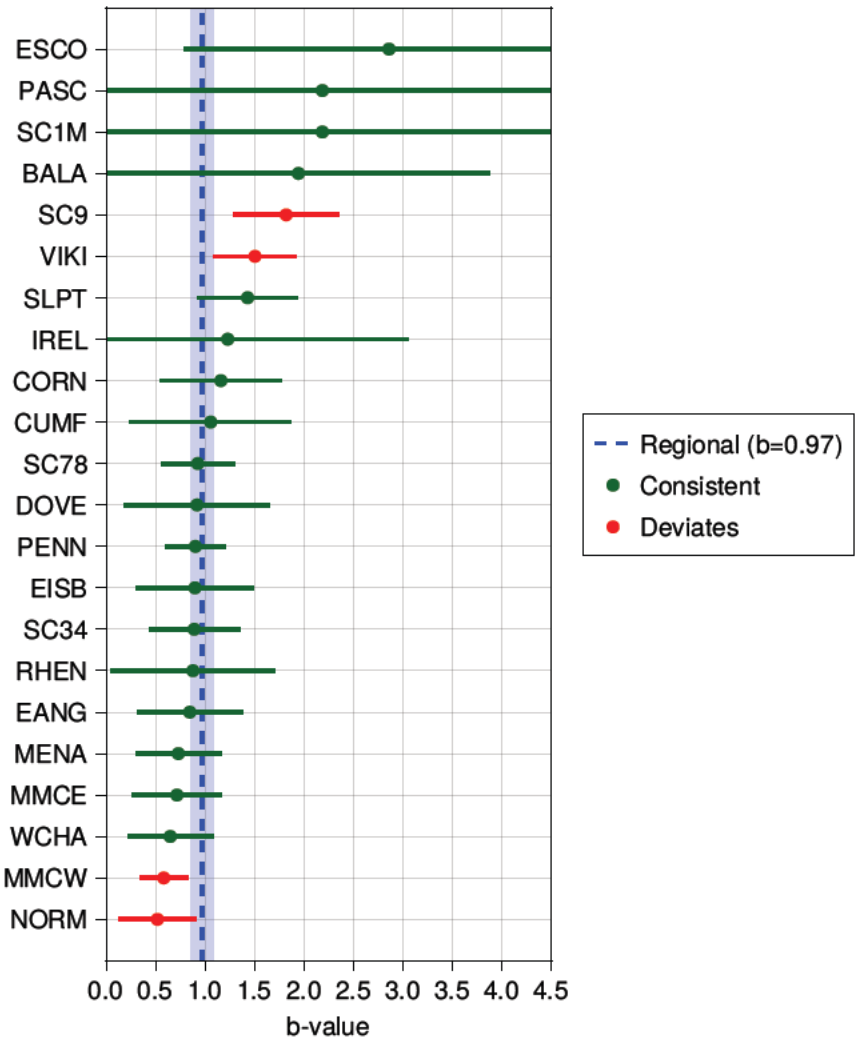

Figure 10.1: Zone-specific MLE b-value estimates with 95% confidence intervals, compared to the regional Weichert MLE estimate (horizontal band). Most zone estimates are consistent with the regional value within their uncertainty bounds, though several peripheral zones (NORM, SC9, VIKI) show notable departures.

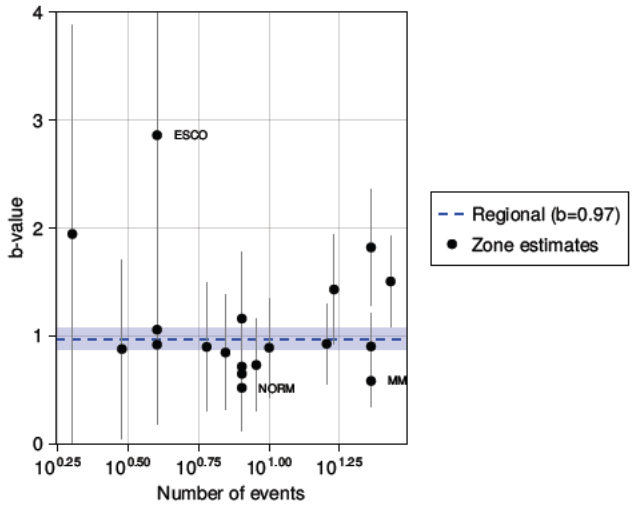

Figure 10.2: Relationship between b-value uncertainty (σ_b) and zone sample size. The funnel shape demonstrates that apparent b-value heterogeneity is predominantly driven by sampling variability in data-sparse zones rather than genuine spatial variation.

10.3.2 Results: MLE Tests

Of 46 adjacent zone pairs, 31 have sufficient data for testing (both zones with at least 2 events above the completeness threshold). The **b-value test** finds only 5 pairs (16%) significantly different at the 5% level. All 5 involve either ESCO ($b = 2.86$ from 4 events) or the SC9–SC78 pair. This is consistent with the hypothesis that all UK zones share a common b-value. The pairwise test results are visualised as a matrix in Figure 10.3, and the zone-by-zone comparison to the regional estimate is shown in Figure 10.4.

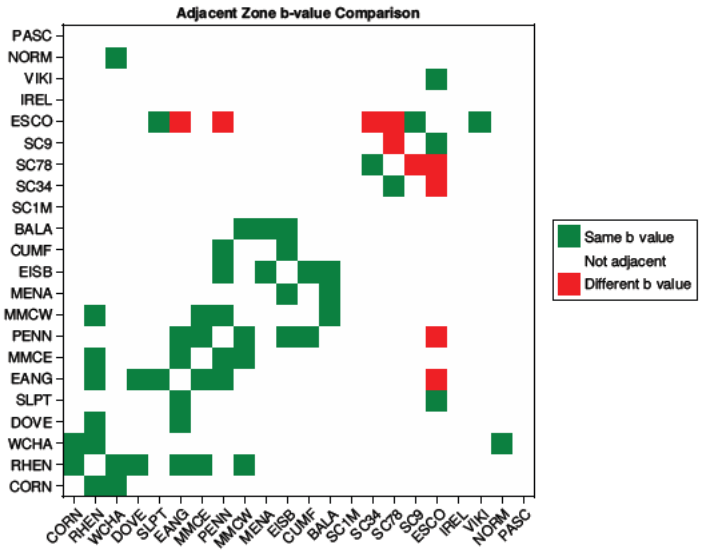

Figure 10.3: Likelihood ratio test results for adjacent zone pairs testing the null hypothesis of shared b -value. Red cells indicate statistically significant differences ($p < 0.05$); blue cells indicate no significant difference. Grey cells denote non-adjacent or untestable pairs. Only 5 of 31 testable pairs show significant b-value differences.

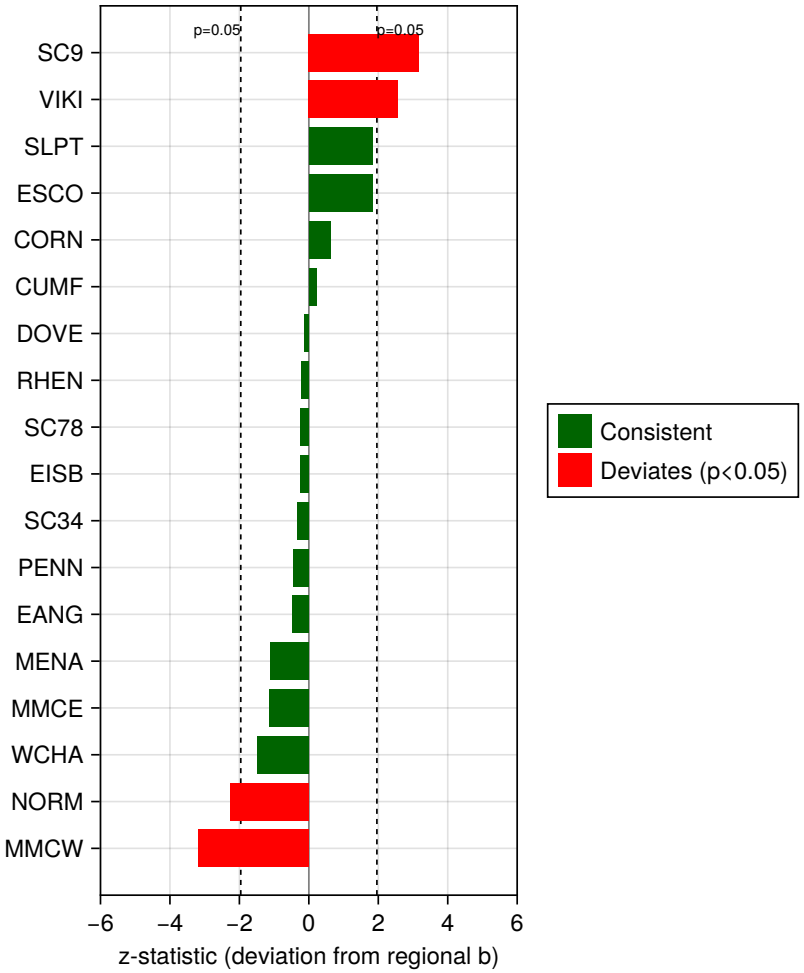

Figure 10.4: Zone-specific z-statistics comparing each zone's MLE b-value to the regional Weichert estimate ($b = 0.97$). Horizontal dashed lines mark the ± 1.96 significance thresholds. Zones outside these bounds have b-values that differ significantly from the national average at the 5% level.

The **full aggregation test**, which additionally requires consistent area-normalised activity rates, gives a substantially different picture: 21 of the 31 testable pairs (68%) are significantly different. The discrepancy is driven by 16 zone pairs that have indistinguishable b-values but significantly different seismicity densities. For example, EANG and PENN have virtually identical b-values (0.85 vs 0.90, $p = 0.86$) but very different area-normalised rates (0.01 vs 0.12 events per 10^4 km² per year, $p < 0.001$). Similarly, MENA has a seismicity density of 0.40 events per 10^4 km² per year — an order of magnitude higher than adjacent zones — reflecting its small area and relatively high activity.

10.3.3 Results: Effect of the PMLM Prior

The same tests were repeated using PMLM fitting (with the Johnston prior $b \sim \mathcal{N}(1.0, 0.087^2)$) to investigate whether the prior-induced shrinkage of b-values and the associated adjustment of activity rates (through the positive $\rho(\lambda_n, \beta)$ correlation) affect the conclusions.

Table 10.3: Comparison of likelihood ratio tests using MLE and PMLM fitting. The PMLM prior ($\sigma_b = 0.087$) completely homogenises b-value estimates but has negligible effect on the full aggregation test, demonstrating that zone distinguishability is driven by activity rate differences rather than b-value differences.

Test	Distinguishable	Indistinguishable
MLE b-value (df=1)	5 / 31 (16%)	26 / 31 (84%)
MLE full aggregation (df=2)	21 / 31 (68%)	10 / 31 (32%)
PMLM b-value (df=1)	0 / 31 (0%)	31 / 31 (100%)
PMLM full aggregation (df=2)	20 / 31 (65%)	11 / 31 (35%)

The PMLM prior has a dramatic effect on b-value estimates — compressing the MLE range of $b = 0.52$–2.86 into a PMLM range of $b = 0.88$–1.11 — with the consequence that no zone pairs show significantly different b-values. However, the full aggregation test is barely affected: 20 pairs remain distinguishable under PMLM compared to 21 under MLE. Only one pair (SC78–SC9, where the MLE b-value difference of 0.93 vs 1.82 was the main driver) switches from "different" to "mergeable" when the prior pulls both b-values toward 1.0.

This confirms that the prior-induced positive correlation between λ_n and β does not materially alter the activity rate picture. The b-values are pulled together, and the λ_n estimates shift correspondingly, but the area-normalised rate differences between zones are large enough to remain significant regardless.

The 10–11 zone pairs that pass the full aggregation test are concentrated among the low-activity zones of central and eastern England (RHEN, WCHA, EANG, MMCE, NORM) and the northern Pennine/Cumbrian region (PENN, CUMF, EISB, BALA):

Table 10.4: Adjacent zone pairs that are not statistically distinguishable under the full aggregation test. ρ_i denotes area-normalised activity rate (events per 10^4 km^2 per year). The SC78–SC9 pair is only mergeable under PMLM, where the prior removes the b-value difference.

Pair	ρ_1	ρ_2	p (MLE)	p (PMLM)
RHEN – WCHA	0.01	0.01	0.69	0.81
RHEN – EANG	0.01	0.01	0.65	0.66
RHEN – MMCE	0.01	0.02	0.19	0.21
WCHA – NORM	0.01	0.01	0.46	0.69
DOVE – EANG	0.07	0.01	0.06	0.06
EANG – MMCE	0.01	0.02	0.57	0.64
PENN – MMCW	0.12	0.07	0.20	0.91
PENN – CUMF	0.12	0.07	0.42	0.45
EISB – CUMF	0.04	0.07	0.79	0.83
EISB – BALA	0.04	0.04	0.35	0.72
SC78 – SC9	0.22	0.27	0.014	0.54

10.3.4 Sensitivity to Minimum Magnitude Threshold

The preceding results use $M_w \geq 3.0$ as the minimum magnitude, which limits the analysis to 218 completeness-filtered events and leaves 15 of the 46 adjacent zone pairs untestable due to data sparsity. Lowering the threshold to $M_w \geq 2.5$ (complete from 1970) roughly doubles the catalogue to 426 events and, crucially, ensures that all 22 zones have at least 2 events — making all 46 adjacent pairs testable.

Table 10.5: Effect of minimum magnitude threshold on zone distinguishability test results. "pp" denotes percentage points. The PMLM full aggregation test — the most relevant test for source model structure — gives identical proportions at both thresholds.

Test	$M_w \geq 3.0$	$M_w \geq 2.5$	Change
MLE b-value (df=1)	5 / 31 (16%)	17 / 46 (37%)	+21 pp
MLE full aggregation (df=2)	21 / 31 (68%)	36 / 46 (78%)	+10 pp
PMLM b-value (df=1)	0 / 31 (0%)	1 / 46 (2%)	+2 pp
PMLM full aggregation (df=2)	20 / 31 (65%)	30 / 46 (65%)	0 pp

The results reinforce the conclusions from the primary analysis:

- **MLE b-value differences increase substantially** (16% to 37%). With more data, genuine or apparent b-value heterogeneity that was hidden by sampling noise at $M_w \geq 3.0$ becomes statistically detectable. However, most of these differences involve zones where the MLE b-value is poorly constrained (e.g. RHEN rises from 3 to 17 events, but its b-value estimate of 1.44 is still driven by the predominance of small events in a limited sample).
- **The PMLM prior continues to homogenise b-values almost completely** — only 1 of 46 pairs (2%) shows a significant difference, compared to 0 of 31 at $M_w \geq 3.0$. The Johnston prior is equally effective at pooling b-values regardless of the magnitude threshold.
- **The full aggregation test is remarkably stable.** Under PMLM, exactly 65% of testable pairs are distinguishable at both thresholds. The additional 15 pairs that become testable at $M_w \geq 2.5$ split in roughly the same proportions as the original 31. This stability confirms that the conclusions about zone distinguishability are not an artifact of the particular magnitude threshold chosen.
- **More MLE–PMLM disagreements emerge** — 6 pairs at $M_w \geq 2.5$ versus 1 at $M_w \geq 3.0$. In all cases, the MLE test flags a difference (driven by b-value) that the PMLM prior suppresses. This is consistent with the interpretation that the additional b-value heterogeneity detected by MLE at lower thresholds is largely sampling noise that the prior correctly regularises.

The consistency of the PMLM full aggregation results across magnitude thresholds provides strong evidence that the main conclusions — b-values are not individually resolvable while

activity rates are — are robust properties of the UK seismicity data rather than threshold-dependent artifacts.

10.3.5 Sensitivity to Boundary Placement

The distinguishability tests treat zone boundaries as fixed. In practice, these boundaries are uncertain — they represent expert judgment about the spatial limits of seismotectonic provinces, and reasonable alternative placements could differ by several kilometres. A natural question is whether small boundary shifts could change the test conclusions for marginally significant pairs.

To address this, a boundary shift analysis was performed for each PMLM-distinguishable pair. For a given shift distance d (in km), the shared boundary is displaced toward the higher-density zone, transferring a strip of area $\Delta A \approx L \times d$ (where L is the shared boundary length) and all events within distance d of the boundary on the higher-density side. The full aggregation test is then recomputed with the adjusted areas and reassigned events. The critical shift d^* is the minimum displacement at which the pair becomes mergeable ($p > 0.05$).

Table 10.6: Boundary shift analysis for PMLM-distinguishable pairs. p is the original PMLM full aggregation p-value, L is the shared boundary length, d^* is the critical shift needed to make the pair mergeable, and "Events moved" is the number of earthquakes reassigned. Only pairs with $d^* \leq 50$ km or $p \geq 0.01$ are shown. Results are categorised as Fragile ($d^* \leq 5$ km), Moderate (5–15 km), Stable (15–50 km), or Robust (d^* exceeds the search limit).

Pair	p	L (km)	d^* (km)	Events moved	Category
PENN – EISB	0.048	110	3.5	1	Fragile
SC34 – SC78	0.009	104	2.0	2	Fragile
CORN – WCHA	0.033	194	9.0	1	Moderate
RHEN – DOVE	0.018	50	6.5	1	Moderate
MENA – BALA	0.000	65	8.0	5	Moderate
RHEN – MMCW	0.000	199	24.5	10	Stable
SLPT – EANG	0.002	323	32.5	7	Stable
MMCE – MMCW	0.001	196	32.0	8	Stable
CORN – RHEN	0.017	161	> 50	—	Robust
MMCW – BALA	0.041	171	> 50	—	Robust

Of 20 PMLM-distinguishable pairs at $M_w \geq 3.0$:

- **2 pairs are fragile** ($d^* \leq 5$ km): PENN–EISB and SC34–SC78. A boundary shift of just 2–3.5 km — well within the uncertainty of expert zonation — would make these pairs statistically indistinguishable. In both cases, only 1–2 events need to cross the boundary.
- **3 pairs are moderate** (5–15 km): CORN–WCHA, RHEN–DOVE, and MENA–BALA. These require non-trivial but not unreasonable boundary adjustments.

- **5 pairs are stable** (15–50 km): These need substantial boundary changes and are unlikely to be affected by reasonable zonation uncertainty.
- **10 pairs are robust** ($d^* > 50$ km or physically unreachable): These reflect genuine, large-scale rate contrasts that no reasonable boundary adjustment could eliminate.

The fragile cases are instructive. PENN–EISB ($p = 0.048$) is already the most marginal rejection in the entire analysis. A single event lies within 3.5 km of the PENN–EISB boundary on the PENN (higher-density) side; reassigning it and the corresponding area strip to EISB makes the pair indistinguishable. SC34–SC78 ($p = 0.009$) has a lower p-value, but the short boundary length means just 2 events within 2 km of the boundary drive the distinction; their reassignment to the lower-density zone eliminates the significance.

At $M_w \geq 2.5$, only 1 pair is fragile: DOVE–EANG ($d^* = 2$ km, 1 event). The two pairs that were fragile at $M_w \geq 3.0$ become moderate (PENN–EISB, $d^* = 10.5$ km) or are no longer in the "different" category at this threshold. The additional data generally stabilise the boundary conclusions, with 21 of 30 distinguishable pairs classified as robust.

10.3.5.1 Effect of epicentral location uncertainty

The boundary shift analysis assumes that event locations are known exactly. In practice, epicentral locations carry uncertainties that are comparable to the critical shift distances. The BGS catalogue assigns quality factors to historical (pre-instrumental) events based on macroseismic data: A denotes maximum epicentral error < 5 km, B denotes 5–15 km, C denotes 15–30 km, and D denotes > 30 km. Instrumental events (post-1970) have typical lateral uncertainties of approximately 3 km.

Examination of the specific events near fragile and moderate boundaries reveals that location uncertainty is directly relevant:

- **PENN–EISB (fragile, $d^* = 3.5$ km)**: The critical event is a 1777 earthquake ($M_w = 4.1$) lying just 3.3 km from the boundary, with quality factor Ba (5–15 km maximum error). Its location uncertainty exceeds the critical shift distance. Of the 29 events in these two zones, 15 (52%) are historical. The mean location uncertainty is 8.0 km.
- **SC34–SC78 (fragile, $d^* = 2.0$ km)**: The two critical events are a 1907 historical earthquake ($M_w = 4.1$, quality Ba, 0.5 km from boundary) and a 1998 instrumental event ($M_w = 3.2$, 1.8 km from boundary). The 1907 event has a location uncertainty of 5–15 km — an order of magnitude larger than its reported distance to the boundary. Of 26 events, 15 (58%) are historical.
- **RHEN–DOVE (moderate, $d^* = 6.5$ km)**: Only 7 events total (3 historical, 4 instrumental), making this pair highly sensitive to any single event moving between zones.

To quantify the impact of location uncertainty, a Monte Carlo analysis was performed. In each of 2000 realisations, all event epicentres were perturbed by Gaussian noise with standard

deviation determined by their quality factor ($\sigma = 5$ km for A, 10 km for B, 22.5 km for C, 30 km for D) or 3 km for instrumental events. These values are deliberately conservative (if the BGS uncertainty assessments are taken as the truth): the BGS quality factor documentation specifies maximum errors of < 5 km (A), 5–15 km (B), 15–30 km (C), and > 30 km (D), and the adopted σ values treat these limits as approximate 1σ uncertainties rather than bounding values. Events were then reassigned to zones based on their perturbed locations, and the PMLM full aggregation test was recomputed.

Table 10.7: Effect of epicentral location uncertainty on zone distinguishability. "Different" and "Mergeable" show the percentage of 2000 Monte Carlo realisations classified as statistically distinguishable or mergeable under the PMLM full aggregation test after perturbing event locations. "Historical" is the number of pre-1970 events out of the total in both zones.

Pair	p	Different	Mergeable	Historical	Mean σ (km)
PENN – EISB	0.048	12%	88%	15 / 29	8.0
SC34 – SC78	0.009	58%	42%	15 / 26	6.4
RHEN – DOVE	0.018	66%	34%	3 / 7	6.0
CORN – WCHA	0.033	87%	13%	11 / 16	13.6
MENA – BALA	0.000	99%	1%	4 / 11	5.5

The results confirm and strengthen the fragility classification. PENN–EISB is classified as mergeable in 88% of location realisations — once epicentral uncertainty is properly accounted for, this zone pair is more likely indistinguishable than not. The test outcome is essentially determined by the uncertain location of a single 18th-century earthquake. SC34–SC78, which appeared clearly distinct ($p = 0.009$) under the deterministic analysis, becomes mergeable in 42% of realisations — nearly a coin-flip once location uncertainty is acknowledged. RHEN–DOVE is also moderately sensitive (34% mergeable), reflecting the very small sample size (7 events total) and the influence of individual event placements. CORN–WCHA, classified as stable under the boundary shift analysis, nonetheless becomes mergeable in 13% of realisations, driven by its high proportion of historical events (11 of 16) with substantial location uncertainties.

Only MENA–BALA is genuinely robust to location perturbation, remaining distinguishable in 99% of realisations despite having 4 historical events among its 11 total. This reflects the large rate contrast between these zones, which is too great to be bridged by plausible epicentral shifts. These results demonstrate that most zone boundaries are well-supported by the data, but a significant minority — particularly those depending on historical events near boundary segments — are sensitive to location uncertainty that is routinely neglected in source model construction.

10.3.6 Results: Comparison to Regional b-value

Of 18 zones with sufficient data for reliable estimation (at least 3 events with a converged MLE fit):

- **14 zones** (78%) are consistent with the regional b-value ($p > 0.05$)
- **4 zones** deviate significantly:

Table 10.8: Source zones whose MLE b-value deviates significantly ($p < 0.05$) from the Weichert MLE regional estimate of $b = 0.97$ (Table 10.2). In most cases the deviation is attributable to data artifacts rather than genuine spatial variation in seismicity.

Zone	b	p-value	Likely explanation
MMCW	0.58	0.001	Bipartite FMD artifact
SC9	1.82	0.001	Possibly genuine / offshore effects
VIKI	1.51	0.010	Offshore zone
NORM	0.52	0.023	Channel zone / completeness

MMCW exhibits a bipartite frequency-magnitude distribution, so fitting a single GR model produces a spuriously low b-value that represents neither the upper nor lower branch. SC9 and VIKI are offshore zones where the completeness model may not accurately reflect the higher detection thresholds. NORM lies in the English Channel where similar completeness concerns apply.

10.3.7 Implications for Source Model Structure

The tests answer two distinct questions, and the PMLM comparison clarifies which aspects of the source model are well-constrained by the data:

B-values are not individually resolvable. Under PMLM, 0% of adjacent pairs show significantly different b-values at $M_w \geq 3.0$, and only 2% at $M_w \geq 2.5$. Even under pure MLE (without any prior), the proportion is only 16% at $M_w \geq 3.0$ — and while this rises to 37% at $M_w \geq 2.5$ as more data reveals apparent heterogeneity, the PMLM prior correctly regularises these differences (Table 10.5). This strongly supports pooling b-values across zones — either through a shared prior (as in current PMLM practice) or through a hierarchical model that estimates the between-zone variability from the data.

This finding has a practical consequence for logic tree computation. In principle, rigorous propagation of epistemic uncertainty through N source zones with $n_b \times n_\lambda$ branches each requires enumerating all $(n_b \times n_\lambda)^N$ joint parameter combinations — $9^{22} \approx 10^{21}$ end branches for a 3×3 scheme across 22 zones, or $25^{22} \approx 6 \times 10^{30}$ for 5×5. If b-values are shared, this reduces to $n_b \times n_\lambda^N$ ($3^{23} \approx 10^{11}$ or $5^{23} \approx 10^{16}$) — reductions by factors of 10^{10} and 5×10^{14} respectively.

In practice, however, hazard calculations rarely attempt this full factorial enumeration. The standard approach treats each source zone independently: hazard contributions are

computed per-source for each branch and then summed, with the mean hazard being exact under this decomposition. For mean hazard, a shared b-value offers no computational saving — each source still requires the same number of per-source hazard evaluations regardless of whether its b-value branches are independent or shared with other zones.

The real significance of a shared b-value is not computational but structural: it correctly represents the epistemic correlation between sources. If the data cannot distinguish b-values across zones, then the epistemic uncertainty in b is a common (shared) uncertainty, and b-values should be perfectly correlated across sources in the logic tree — when one source's b-value is high, all sources' b-values should be high. This correlation has no effect on mean hazard but directly affects fractile hazard estimates, which are the quantities that can drive regulatory decisions. Computing these fractiles correctly *does* require tracking joint parameter combinations across sources, and this is precisely where the shared-b structure becomes both computationally advantageous and scientifically necessary: the b-value node sits above the source-specific branching in the logic tree, reducing the dimensionality of the joint enumeration from $2N$ to $N + 1$ independent parameters.

Activity rates are individually resolvable. Approximately two-thirds of adjacent zone pairs have significantly different area-normalised seismicity densities under the full aggregation test. This proportion is remarkably stable: 65% under PMLM at both $M_w \geq 3.0$ and $M_w \geq 2.5$ (Table 10.5), and robust to the choice of fitting method (MLE or PMLM). The prior-induced correlation between λ_n and β does not materially homogenise the rates. The existing zone boundaries therefore do delineate regions with genuinely different seismicity characteristics — but the differences are driven by activity rate, not by b-value.

The PMLM prior is effective for b-value pooling but irrelevant for rate estimation. The Johnston prior successfully regularises b-value estimates without distorting the rate picture, and does so consistently across magnitude thresholds. This is a useful practical finding: practitioners can apply the PMLM prior with confidence that it will not artificially homogenise the activity rates that ultimately drive hazard at most sites.

Most zone boundaries are robust to plausible perturbations. The boundary shift analysis (Table 10.6) shows that 75% of distinguishable boundaries (15 of 20 at $M_w \geq 3.0$) require shifts exceeding 15 km to change the test outcome — well beyond reasonable zonation uncertainty. Only 2 pairs (PENN–EISB and SC34–SC78) are fragile, where shifts of 2–3.5 km and the reassignment of 1–2 events would make them indistinguishable. These represent boundaries where the current zonation is at the limit of what the data can resolve, and where future model revisions should consider alternative placements.

These findings directly motivate the hierarchical modelling approach explored in Section 12: zones should share a common b-value distribution (reducing the effective number

of b-value parameters), while retaining zone-specific activity rates to capture real spatial variation in seismicity density.

10.4 Correlation Structure in Source Modelling

The preceding analysis established that UK source zone b-values are not individually resolvable and that a shared b-value — whether imposed via a common prior or estimated through a hierarchical model — is the natural description of the data. This has a direct consequence that standard PSHA practice does not currently address: if the epistemic uncertainty in b is genuinely shared across zones, then treating zone parameters as independent in the logic tree misrepresents the correlation structure of the uncertainty.

Standard practice constructs logic trees with independent branches for each source zone. The total hazard at a site is computed by summing contributions from all sources, and the logic tree weights are applied per-source without cross-zone coupling. For the *mean* hazard, this decomposition is exact — the expected value of a sum is the sum of expectations, regardless of correlation. However, for *fractile* hazard estimates — which can determine regulatory design values — the distribution of total hazard depends critically on how parameter uncertainties co-vary across sources. Ignoring correlations is equivalent to assuming that the epistemic uncertainty in each zone's parameters is independent of every other zone's, which is physically implausible when all zones share the same tectonic setting, the same prior information, and the same moment budget.

Three distinct mechanisms induce correlations between source zone parameters. Each is analysed below using Monte Carlo simulation of a simplified UK source model comprising the 8 largest contributors to the national moment budget (MMCW, PENN, MENA, SLPT, IREL, WCHA, SC34, and an aggregated "Other" zone).

10.4.1 Correlations from Hierarchical b-value Priors

10.4.1.1 The mechanism

The PMLM approach applies a common prior $b \sim \mathcal{N}(\mu_b, \sigma_b^2)$ to every zone independently. Each zone's posterior b-value is therefore a weighted average of its zone-specific data and the shared prior mean — the classic "partial pooling" or "shrinkage" behaviour. For data-sparse zones, the posterior is dominated by the prior and all such zones converge toward μ_b; for data-rich zones, the posterior is dominated by the data and the prior has less influence. The net effect is that zone b-values are *more similar to each other* than the data alone would suggest.

However, current practice does not represent this coupling in the logic tree. Each zone is assigned its own b-value branches and weights, sampled independently. This is internally

inconsistent: the PMLM prior imposes a belief that all zones share a common regional b-value, but the logic tree then treats the resulting estimates as if they were unrelated.

10.4.1.2 Hierarchical model formulation

To quantify the induced correlation, a two-level hierarchical model was constructed. At the upper level, a regional b-value is drawn from a prior representing genuine uncertainty about the UK-wide b:

$$b_{\text{regional}} \sim \mathcal{N}\left(\mu_b, \sigma^2_{\text{regional}}\right) \tag{10.6}$$

At the lower level, each zone's b-value is drawn as a deviation from the regional value, representing zone-to-zone variability:

$$b_i \mid b_{\text{regional}} \sim \mathcal{N}\left(b_{\text{regional}}, \sigma^2_{\text{zone}}\right) \quad i = 1, ..., N \tag{10.7}$$

Activity rates are sampled independently for each zone:

$$\log_{10} \lambda_{n,i} \sim \mathcal{N}\left(\mu_{\lambda,i}, \sigma^2_{\lambda,i}\right) \tag{10.8}$$

with zone-specific means and standard deviations taken from the NSHM 2020 estimates.

The key distinction from standard PMLM is that the hierarchical model explicitly represents b_{regional} as a shared latent variable. In any single Monte Carlo realisation, all zones share the same drawn value of b_{regional}; their individual b-values differ only by the zone-level perturbation σ_{zone}. This creates a positive correlation between zone b-values whose magnitude depends on the ratio of the two variance components:

$$\rho\left(b_i, b_j\right) = \frac{\sigma^2_{\text{regional}}}{\sigma^2_{\text{regional}} + \sigma^2_{\text{zone}}} \quad \text{for } i \neq j \tag{10.9}$$

When $\sigma_{\text{regional}} \gg \sigma_{\text{zone}}$, the regional uncertainty dominates and all zones move together ($\rho \to 1$). When $\sigma_{\text{zone}} \gg \sigma_{\text{regional}}$, zone-specific variation overwhelms the common signal ($\rho \to 0$).

10.4.1.3 Monte Carlo analysis

The correlation structure was evaluated by drawing 50,000 samples from the hierarchical model and computing the Pearson correlation matrix of the resulting zone b-values. The analysis was repeated across a range of σ_{regional} values (0.01 to 0.20 in b-units) with σ_{zone} fixed at 0.05, reflecting modest but non-zero zone-to-zone variability.

Table 10.9 shows the results. The standard PMLM specification corresponds to a total prior standard deviation $\sigma_b = 0.087$ (i.e. $\sigma_\beta = \frac{1}{\sqrt{W}} = 0.20$ with weight $W = 25$, converted to the b -scale via $\sigma_b = \frac{\sigma_\beta}{\ln(10)}$). If the zone-to-zone component is $\sigma_{\text{zone}} = 0.05$, the implied regional component is $\sigma_{\text{regional}} = \sqrt{0.087^2 - 0.05^2} \approx 0.07$, giving $\rho\left(b_i, b_j\right) \approx 0.67$ from Equation 10.9. If instead σ_{zone} is smaller (say 0.03, reflecting the finding that zone b-values are essentially indistinguishable), the correlation increases to $\rho \approx 0.85$ or higher.

Table 10.9: Induced correlation between zone b-values as a function of the hierarchical variance partition. σ_{regional} is the standard deviation of the shared regional b-value; σ_{zone} is the standard deviation of zone-specific deviations. Correlations computed from Equation 10.9 and confirmed by Monte Carlo simulation with 50,000 samples.

σ_{regional}	σ_{zone}	$\rho(b_i, b_j)$	Interpretation
0.01	0.05	≈ 0.04	Regional b nearly fixed; zone scatter dominates
0.05	0.05	≈ 0.50	Equal regional and zone components
0.07	0.05	≈ 0.67	Consistent with PMLM ($\sigma_b = 0.087$)
0.10	0.05	≈ 0.80	Strong regional uncertainty
0.15	0.05	≈ 0.90	Dominant regional uncertainty
0.20	0.05	≈ 0.94	Zone variation negligible relative to regional

The precise value of ρ depends on how the total PMLM prior variance is partitioned between regional and zone-level components — a decomposition that the PMLM framework itself does not make explicit. However, for any plausible partition, the correlation is substantial ($\rho > 0.5$), and the zone distinguishability results of Section 10 — which found 0% of adjacent pairs with significantly different b-values under PMLM — suggest that σ_{zone} is small, pushing the correlation toward the upper end of the range.

Implication: Treating b-values as independent in the logic tree overstates the total epistemic uncertainty. The effective number of independent b-value degrees of freedom across N zones is closer to 1 (a single shared b) than to N (fully independent).

10.4.2 Correlations from Moment Release Constraints

10.4.2.1 The mechanism

A physical constraint on the total seismic moment release rate creates *negative* correlations between zones. If the total moment release across all zones is constrained to match an external estimate — for example, from geodetic strain rates or glacio-isostatic rebound modelling (see Section 11) — then a higher-than-expected moment release from one zone requires lower release from others to satisfy the budget. This is the opposite sign from the hierarchical b-value effect.

10.4.2.2 Implementation

The moment constraint was applied via importance sampling. First, 50,000 unconstrained samples were drawn from the hierarchical model. For each sample k, the total moment release across all zones was computed as $\dot{M}_{\text{total}}^{(k)} = \sum_i \dot{M}_{0,i}^{(k)}$, where each zone's moment release was calculated from the standard integral over the truncated Gutenberg-Richter distribution with the sampled $\lambda_{n,i}$, b_i, and $M_{\max} = 6.63$ (the weighted mean of the NSHM 2020 $M_{\max}$ logic tree: 6.5 at 50%, 6.7 at 20%, 6.9 at 20%, 7.1 at 10%). Each sample was then assigned an importance weight:

$$w_k \propto \exp\left(-\frac{1}{2}\left(\frac{\dot{M}_{\text{total}}^{(k)} - \dot{M}_{\text{target}}}{\sigma_{\text{target}}}\right)^2\right) \tag{10.10}$$

with $\dot{M}_{\text{target}} = 1.42 \times 10^{16}$ N·m/yr and $\sigma_{\text{target}} = 0.3 \times 10^{16}$ N·m/yr, based on the geodetic estimate of (Main et al., 1999). Weighted correlation matrices were then computed from the reweighted samples. The effective sample size ($\text{ESS} = \frac{1}{\sum_k w_k^2}$) was monitored to ensure adequate sampling.

10.4.2.3 Results

The strength of the moment-constraint-induced correlation depends on each zone's fractional contribution to the total moment budget. Zones that dominate the budget are most strongly coupled:

Table 10.10: Dominant source zones by seismic moment release contribution in the UK NSHM 2020 and the sign of moment-constraint-induced correlations with other zones.

Zone	% of Total $\dot{M}_0$	Correlation with others
MMCW	32%	Strongly negative
PENN	12%	Moderately negative
MENA	10%	Moderately negative
Other zones	$<$ 7% each	Weakly negative to positive

The moment constraint on its own induces mean pairwise correlations of approximately $\rho \approx -0.25$ to -0.5 between the largest contributors (MMCW, PENN, MENA), but weaker correlations ($|\rho| < 0.1$) between the many small zones that each contribute less than 7% of the total moment budget.

10.4.3 Correlations from Boundary Uncertainty

Uncertainty in source zone boundaries creates local negative correlations between adjacent zones. When a boundary shifts, one zone gains area and events while the other loses them. This anti-correlated perturbation to both zone parameters arises simultaneously from a single source of uncertainty (the boundary position). The boundary shift analysis of Table 10.6 quantifies this for specific zone pairs: for fragile boundaries ($d^* \leq 5$ km), the reassignment of even one or two events is sufficient to change the statistical classification, implying that the activity rate estimates of the two zones are strongly anti-correlated conditional on boundary position.

This effect is inherently local — it couples only adjacent zones that share a boundary — and its magnitude depends on the event density near the boundary and the boundary uncertainty. Unlike the first two mechanisms, it cannot be captured by a global hierarchical model and instead requires explicit treatment of boundary geometry.

10.4.4 Combined Correlation Structure

The three correlation sources act simultaneously with different signs and spatial structures:

- **Hierarchical prior**: positive correlation across *all* zone pairs (shared regional b pulls estimates together)
- **Moment constraint**: negative correlation, strongest between zones that dominate the moment budget
- **Boundary uncertainty**: negative correlation, limited to *adjacent* zone pairs with uncertain boundaries

For a given pair of zones, the net correlation is approximately:

$$\rho(\theta_i, \theta_j) \approx \rho_{\text{prior}} + \rho_{\text{moment}} + \rho_{\text{boundary}} \tag{10.11}$$

The dominant term depends on the specific pair. Small, non-adjacent zones that contribute modestly to the total moment budget are governed by the positive prior correlation. Dominant moment-contributing zones (MMCW, PENN, MENA) experience competing effects, with the net sign depending on the relative strength of the hierarchical coupling and the moment constraint. Adjacent zones with fragile boundaries have an additional negative contribution that may dominate locally.

To assess the net effect, the hierarchical sampling was repeated with and without the moment constraint, and the difference in correlation matrices was computed. For a representative configuration ($\sigma_{\text{regional}} = 0.10$, $\sigma_{\text{zone}} = 0.05$), the hierarchy alone produces a mean pairwise correlation of $\rho \approx +0.25$ in log moment release. Adding the moment constraint reduces this to $\rho \approx +0.10$, implying a mean constraint effect of $\Delta\rho \approx -0.15$. The positive hierarchical effect dominates for most zone pairs, but for the largest contributors the moment constraint can produce net negative correlations. The moment release framework is developed fully in Section 11, and the hierarchical model that jointly estimates shared and zone-specific parameters — thereby capturing these correlations explicitly — is presented in Section 12.

10.5 Assessment of PMLM Prior Assumptions for UK Source Zones

The preceding sections have established three findings with direct implications for how the PMLM prior should be understood: (i) zone b-values are not individually resolvable (Section 10.3), suggesting that the prior is providing most of the information about b in data-sparse zones; (ii) a shared regional b-value is supported by the data, which means the prior's role is not merely to regularise noisy estimates but to define the effective b-value for most zones; and (iii) the shared prior induces positive correlations between zone b-values (Section 10.4) that current practice ignores. These findings motivate a systematic examination of whether the specific prior adopted in UK practice — the Johnston PMLM prior

with $\mu_b = 1.0$ and $\sigma_b = 0.087$ — is actually appropriate for the UK, and whether it accurately represents the epistemic uncertainty at the zone level.

10.5.1 The Hierarchical Misspecification Problem

The PMLM framework of Johnston et al. (1994) provides a general method for penalised maximum likelihood estimation but does not prescribe specific prior hyperparameters. The prior mean $\mu_b = 1.0$ reflects the long-standing empirical observation, accumulated over many decades of seismological research, that b-values tend to cluster near unity across diverse tectonic settings. The prior weight $W = 25$ (corresponding to $\sigma_b = 0.087$) adopted in UK practice is generally attributed to Musson (2011), who judged this value to be reasonable for UK applications. Neither the prior mean nor the weight derives from a formal statistical calibration — they are practitioner choices informed by experience and judgment.

The question addressed here is not whether these choices were unreasonable *a priori*, but whether they are consistent with the UK data *a posteriori*, and — more fundamentally — whether the PMLM framework itself captures the right uncertainty structure. Regardless of how the prior hyperparameters were chosen, applying the same prior independently to each zone implicitly embeds a hierarchical assumption about the data-generating process. A proper hierarchical model would distinguish multiple levels of uncertainty:

> **Implied Hierarchical Structure**
>
> **Level 1 (Regional)**: $b_{\text{UK}} \sim \mathcal{N}\left(\mu_b, \sigma^2_{\text{regional}}\right)$ — uncertainty in the UK-wide b-value
>
> **Level 2 (Zone)**: $b_{\text{zone}} \mid b_{\text{UK}} \sim \mathcal{N}(b_{\text{UK}}, \sigma^2_{\text{between}})$ — zone-to-zone variability within the UK
>
> **Level 3 (Data)**: Observed counts given b_{zone} — statistical sampling uncertainty

The total uncertainty on any individual zone's b-value should be the quadrature sum of all applicable levels:

$$\sigma^2_{\text{zone,total}} = \sigma^2_{\text{regional}} + \sigma^2_{\text{between}} + \sigma^2_{\text{sampling}} \tag{10.12}$$

where σ_{regional} is the uncertainty in the UK regional b-value itself, σ_{between} is the true zone-to-zone variability within the UK, and σ_{sampling} is the statistical uncertainty from the finite catalogue in that zone.

The PMLM approach does not make this decomposition explicit — all sources of uncertainty are lumped into a single prior width $\sigma_b = 0.087$. The framework itself does not preclude any component; the prior could in principle be interpreted as encompassing regional, between-zone, and sampling contributions simultaneously. The issue is how it is used in practice. Practitioners treat the PMLM posterior — which is dominated by the prior, giving $\sigma_b^{\text{post}} \approx 0.087$ regardless of how many events are observed — as the total zone-level uncertainty. Whether this value is too large, too small, or about right depends on how it compares

to the true $\sigma_{\text{zone,total}}$. If σ_{between} is small (as the zone distinguishability tests suggest), the prior may be wider than necessary for representing zone-to-zone differences, which is benign for individual zone estimates. However, practitioners do not separately represent the regional component σ_{regional} — the uncertainty about where the UK as a whole sits relative to the assumed prior mean. Even with perfect knowledge of zone-to-zone differences, uncertainty in the shared regional b-value propagates to every zone and should be part of the total epistemic uncertainty. Because the PMLM framework lumps everything into a single parameter, there is no mechanism to ensure that the regional and between-zone components are both adequately represented, and in practice they are not.

10.5.2 Diagnostic Framework

To assess these concerns quantitatively, four diagnostic tests were applied to each of the 17 UK source zones with sufficient data for parameter estimation. The diagnostics compare the PMLM posterior against two benchmarks: the data-driven Weichert MLE (which uses no prior information) and a hierarchical uncertainty estimate constructed from the variance decomposition above.

10.5.2.1 D1: Prior-posterior divergence

This diagnostic quantifies how much the zone's data "surprised" the prior. For Gaussian prior $\mathcal{N}(\mu_0, \sigma_0^2)$ and approximate Gaussian posterior $\mathcal{N}(\mu_1, \sigma_1^2)$, four metrics are computed:

- **KL divergence**: The information gained from the data, computed as

$$D_{\text{KL}} = \ln\left(\frac{\sigma_0}{\sigma_1}\right) + \frac{\sigma_1^2 + (\mu_1 - \mu_0)^2}{2\sigma_0^2} - \frac{1}{2} \tag{10.13}$$

  Values exceeding 2.0 indicate that the data are substantially inconsistent with the prior.

- **Overlap coefficient**: The area of overlap between prior and posterior densities, computed via the Bhattacharyya coefficient. Values below 0.3 indicate poor prior specification.

- **Z-shift**: The displacement of the posterior mean from the prior mean in units of the prior standard deviation,

$$Z = \frac{\mu_1 - \mu_0}{\sigma_0} \tag{10.14}$$

  Values $|Z| > 2$ indicate strong data-prior disagreement.

- **Shrinkage**: The fraction of the posterior estimate attributable to the prior rather than the data,

$$\kappa = \frac{\sigma_{\text{prior}}^{-2}}{\sigma_{\text{prior}}^{-2} + \sigma_{\text{data}}^{-2}} \tag{10.15}$$

  ranging from 0 (entirely data-driven) to 1 (entirely prior-driven). High shrinkage ($\kappa > 0.7$) indicates a prior-dominated estimate.

10.5.2.2 D2: Prior predictive checking

This diagnostic tests whether the observed data are plausible under the prior model. For each zone, 5,000 synthetic catalogues are generated by:

1. Sampling $b \sim \mathcal{N}(1.0, 0.087^2)$ from the PMLM prior
2. Sampling λ_n from a log-normal distribution centred on the zone's observed rate
3. Generating event counts from $n \sim \text{Poisson}(\lambda_n \cdot T)$ and magnitudes from the truncated Gutenberg-Richter distribution with the sampled b

The distribution of summary statistics (event count, mean magnitude) from the synthetic catalogues is then compared to the observed values. If the observed statistics fall in the tails of the prior predictive distribution ($p < 0.05$), the prior is a poor match to the data for that zone.

10.5.2.3 D3: Uncertainty suppression detection

This diagnostic directly compares the PMLM posterior standard deviation against the hierarchical benchmark for each zone. The hierarchical standard deviation is constructed from the variance decomposition:

$$\sigma_{\text{hierarchical}} = \sqrt{\sigma^2_{\text{regional}} + \sigma^2_{\text{between}} + \sigma^2_{\text{sampling}}} \tag{10.16}$$

where σ_{regional} is the standard error of the pooled UK b-value estimate from Table 10.1, $\sigma_{\text{between}} = 0.05$ is an assumed zone-to-zone variability (consistent with the finding that most adjacent zones are indistinguishable), and σ_{sampling} is the Weichert MLE standard error for that specific zone.

The suppression ratio is then:

$$R = \frac{\sigma_{\text{PMLM}}}{\sigma_{\text{hierarchical}}} \tag{10.17}$$

with classification thresholds: $R < 0.7$ indicates severe suppression (PMLM underestimates uncertainty by more than 30%), $0.7 \leq R < 0.9$ indicates moderate suppression, and $R > 0.9$ is considered acceptable. The key feature is that $\sigma_{\text{hierarchical}}$ varies across zones (increasing as sample size decreases), while $\sigma_{\text{PMLM}} \approx 0.087$ is nearly constant — the PMLM posterior width is insensitive to the amount of available data because the prior dominates.

10.5.2.4 D4: Regional b-value assessment

This diagnostic compares the data-driven UK regional b-value (from the pooled catalogue analysis of Table 10.1) directly with the PMLM prior mean. The Z-score,

$$Z_{\text{regional}} = \frac{\hat{b}_{\text{UK}} - \mu_{\text{prior}}}{\sigma_{\text{prior}}} \tag{10.18}$$

measures whether the UK data are consistent with the assumed prior centre. A large $|Z_{\text{regional}}|$ indicates that the prior mean should be recalibrated for UK applications.

10.5.3 Key Findings

Finding 1: The prior mean is consistent with Weichert estimates but not with the Full Bayesian

The regional b-value estimates depend critically on the analysis method. The Weichert MLE, which does not account for magnitude uncertainty, yields $b = 0.93 \pm 0.05$ at $M_w \geq 3.0$ and $b = 0.96 \pm 0.03$ at $M_w \geq 2.5$ — both within approximately 1σ of the PMLM prior mean of 1.0 ($Z = -0.8$ and $Z = -0.5$ respectively). However, the Full Bayesian L5 model, which properly accounts for magnitude measurement uncertainty, conversion error, and selection effects via the marginalised likelihood (Section 6), yields substantially lower estimates: $b = 0.85 \pm 0.04$ at $M_w \geq 3.0$ ($Z = -1.7$) and $b = 0.86 \pm 0.03$ at $M_w \geq 2.5$ ($Z = -1.6$). These are approximately 0.10–0.12 below the Weichert estimates, reflecting the Eddington bias correction: methods that ignore magnitude uncertainty overestimate b because events scattering upward across the completeness threshold are preferentially included. The Full Bayesian results place the UK regional b-value nearly 2σ below the prior mean, indicating moderate tension with the assumed $b = 1.0$. This suggests that the prior mean may warrant modest downward revision for UK applications — a value of $b \approx 0.90$ would better centre the prior on the data-supported range across both methods.

Finding 2: PMLM systematically underestimates zone-level uncertainty

PMLM posterior uncertainties are essentially constant at $\sigma_b \approx 0.087$ regardless of sample size, while hierarchical uncertainties range from approximately 0.12 (data-rich zones with $N > 20$ events) to 0.45 (data-poor zones with $N < 10$ events). All 17 zones exhibit severe suppression ($R < 0.7$), with ratios ranging from 0.18 to 0.60. The most data-rich zones show approximately 40% suppression, while data-poor zones show up to 80% suppression. This occurs because the PMLM posterior width is dominated by the prior and therefore insensitive to sample size, whereas the hierarchical benchmark correctly widens as data become scarcer (increasing σ_{sampling}) while also including the irreducible regional uncertainty component (σ_{regional}) that practitioners do not separately represent when using PMLM.

Finding 3: The hierarchical structure is being ignored

The PMLM approach treats each zone as an independent draw from a global distribution, but zones within the UK share a common tectonic setting and — as demonstrated in Section 10.4 — should exhibit correlated b-values. By ignoring this hierarchical structure, PMLM as currently used underestimates total epistemic uncertainty because the regional component σ_{regional} is not separately represented: even if every zone's b were perfectly estimated, there would remain genuine uncertainty about the shared UK regional value that is not explicitly captured in the logic tree. This is precisely the uncertainty that would be represented by the b_{regional} node in the hierarchical model of Section 10.4 and that produces the positive correlations between zone b-values.

10.5.4 Recommendations

1. **For mean hazard applications**: PMLM remains appropriate if the primary output is expected hazard, but the prior mean may warrant modest revision. The Weichert MLE gives $b \approx 0.93$–0.96, consistent with the assumed $b = 1.0$, but the Full Bayesian L5 analysis yields $b \approx 0.85$–0.86 after correcting for magnitude uncertainty effects. A prior mean of $b \approx 0.90$ would better centre the prior on the range of defensible estimates. The uncertainty suppression has limited impact on mean hazard because the PMLM point estimates are adequate.
2. **For fractile-sensitive applications**: Either use hierarchical Bayesian models that explicitly estimate between-zone variability and regional uncertainty, or inflate the PMLM prior standard deviation to approximately $\sigma_b = 0.15$–0.20 to account for the missing hierarchical structure. The choice of inflated σ_b should reflect the combined regional and between-zone uncertainty rather than the global between-region value.
3. **For consistency**: Report both PMLM and data-driven (Weichert or full Bayesian) uncertainty estimates to provide transparency about the sensitivity to prior assumptions.
4. **For future studies**: Develop UK-specific priors calibrated from the combined UK catalogue rather than relying on global SCR analogues. The regional b-value analysis of Table 10.1 provides a natural starting point for such calibration.

11 Theme B: Seismic Moment Release Constraints

Standard practice in PSHA derives Gutenberg-Richter parameters for each source zone independently, using only the earthquake data assigned to that zone. When these independently calibrated zones are then assembled into a composite source model — typically with a common $M_{\max}$ distribution applied across all zones in UK applications — the total seismic moment release rate implied by the combined model is not guaranteed to be consistent with external constraints on regional energy release. Geodetic strain measurements, glacio-isostatic rebound modelling, and geological slip rate estimates each provide independent bounds on the rate at which elastic strain energy is being accumulated and released across the UK. A seismicity model whose implied total moment release falls far outside these bounds is physically implausible (or raises questions about how temporally stable the earthquake processes are in the region, and hence how sound the assumption is to assume that past seismicity reflects future seismicity), regardless of how well its individual zone parameters fit the local earthquake catalogue. This is not to say that moment release constraints are without their own uncertainties, but rather that they provide an independent sanity check on the seismicity model.

Checking the implied moment release against external constraints is therefore a basic sanity test that should be applied to any composite source model, and is not without precedent (Stafford et al., 2008). The calculation itself is straightforward: the annual seismic moment release rate from a Gutenberg-Richter source can be computed analytically as an integral of the moment-magnitude relation over the frequency-magnitude distribution (Baker et al., 2021; McGuire, 1995), and the zone contributions summed to obtain a regional total.

Beyond this basic plausibility check, however, the moment constraint also has a more structural role in the treatment of epistemic uncertainty. As previewed in Section 10.4, when the total moment release is constrained, the parameters of individual zones can no longer be treated as independent: a higher-than-expected rate in one zone must be compensated by lower rates elsewhere to satisfy the budget. This induces *negative* correlations between zone parameters — the opposite sign from the positive correlations created by shared b-value priors. The moment constraint can therefore be exploited not only to identify implausible logic tree branches, but also to introduce physically motivated correlations between sources. This section develops the mathematical framework for computing implied moment release, establishes a geodetic target for the UK, and explores both applications: plausibility checking and correlation induction.

11.1 Geodetic Constraints for the UK

Main et al. (1999) provided estimates of both seismic and tectonic moment release rates for the UK based on glacio-isostatic recovery modelling (Table 11.1).

Table 11.1: Seismic and tectonic moment release rate estimates for the UK from Main et al. (1999), based on glacio-isostatic recovery modelling. The seismic efficiency of 34% indicates that approximately two-thirds of tectonic deformation occurs aseismically.

Quantity	Value
Seismic moment release rate $\dot{M}_{\text{seismic}}$	1.42×10^{16} N·m/year
Tectonic moment release rate $\dot{M}_{\text{tectonic}}$	4.14×10^{16} N·m/year
Seismic efficiency γ	34%

11.2 Implied Moment Release from UK NSHM 2020

Using the mathematical framework developed below (Equation 11.3), the expected annual seismic moment release rate implied by the UK NSHM 2020 source model (Mosca et al., 2022) was computed by summing contributions from all 22 source zones, using the central (preferred) GR parameters and the weighted-mean $M_{\max} = 6.63$:

$$E\left[\dot{M}_0\right] = 1.57 \times 10^{16} \text{ N·m/year} \tag{11.1}$$

This is remarkably consistent with the Main et al. (1999) geodetic estimate (1.42×10^{16} N·m/year, Table 11.1), validating that the NSHM 2020 source model — at its central parameter values — yields physically plausible moment release rates. However, as the elasticity analysis below demonstrates, the moment release at extreme logic tree branch combinations can depart from this central value by orders of magnitude, which is where the constraint becomes most useful.

The moment budget is dominated by a small number of zones (Figure 11.1). MMCW alone contributes approximately 32% of the total, with the three largest contributors (MMCW, PENN, MENA) accounting for roughly 50%. This concentration means that the moment constraint primarily constrains the parameters of these dominant zones, while having limited influence on the many smaller zones that each contribute less than 5% of the total budget. The cumulative contribution curve shows that 12 of 22 zones are needed to reach 90% of total moment release — the remaining 10 zones collectively contribute only 10%.

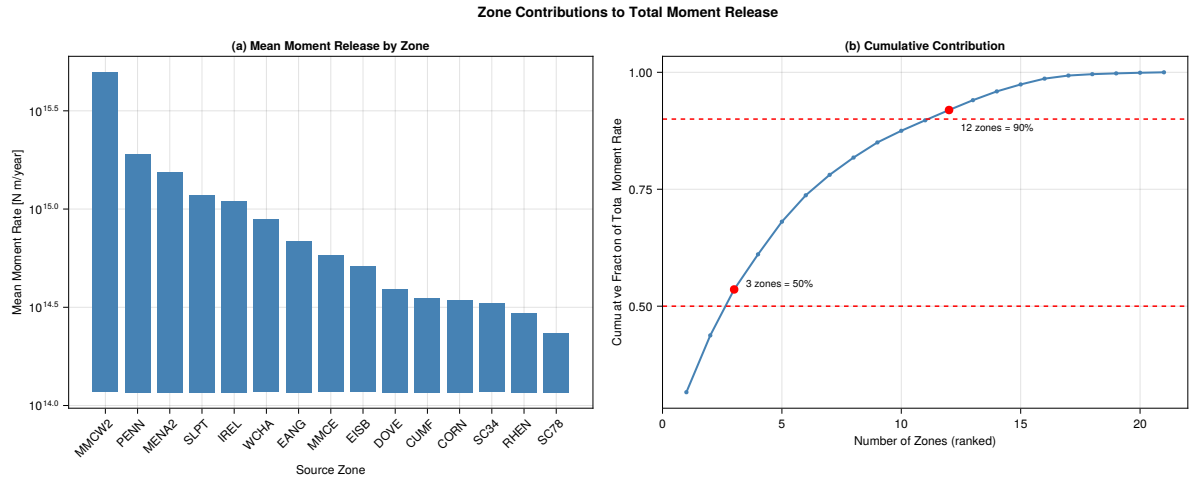

Figure 11.1: Zone contributions to total seismic moment release in the UK NSHM 2020 source model. (a) Mean annual moment release rate by zone, computed from central GR parameters and $M_{\max} = 6.63$. (b) Cumulative fraction of total moment release as zones are added in rank order. Three zones (MMCW, PENN, MENA) account for approximately 50% of the total budget; 12 zones are needed to reach 90%.

11.3 Mathematical Framework

The seismic moment of an earthquake with moment magnitude M_w is given by the Hanks & Kanamori (1979) relation:

$$M_0 = 10^{1.5 M_w + 9.1} \quad [\mathrm{N{\cdot}m}] \tag{11.2}$$

The annual seismic moment release rate from a source characterised by a truncated Gutenberg-Richter distribution is obtained by integrating the product of the moment-magnitude relation and the frequency-magnitude distribution over the admissible magnitude range (Baker et al., 2021). Writing the moment relation in exponential form as $M_0(m) = k e^{cm}$ where $c = 1.5 \ln(10) \approx 3.45$ and $k = 10^{9.1}$, and using the truncated exponential density $f_{M(m)} = \beta e^{-\beta(m - M_{\min})} / \left(1 - e^{-\beta(M_{\max} - M_{\min})}\right)$, the expected annual moment release rate is:

$$\begin{aligned} \dot{M}_0 &= \lambda_n \int_{M_{\min}}^{M_{\max}} M_0(m) f_M(m)\, \mathrm{d}m \\ &= \lambda_n \cdot \frac{\beta \cdot k e^{c M_{\min}}}{(c - \beta)\left(1 - e^{-\beta(M_{\max} - M_{\min})}\right)} \left[e^{(c-\beta)(M_{\max} - M_{\min})} - 1 \right] \end{aligned} \tag{11.3}$$

provided $c \neq \beta$ (i.e. $b \neq 1.5$), which is satisfied for all physically realistic b-values. Here λ_n is the activity rate above $M_{\min}$, and the integral can be evaluated in closed form because both the moment-magnitude relation and the GR density are exponential in m. This expression makes explicit the exponential sensitivity to $M_{\max}$: the term $e^{(c-\beta)(M_{\max} - M_{\min})}$ grows extremely rapidly because $c - \beta > 0$ for any $b < 1.5$.

For a source model comprising N zones, the total annual moment release rate is the sum of zone contributions:

$$\dot{M}_{\text{total}} = \sum_{z=1}^{N} \dot{M}_{0,z}(\lambda_{n,z}, b_z, M_{\max,z}) \tag{11.4}$$

Each zone contributes according to its own GR parameters and maximum magnitude, so the total moment rate is a function of $3N$ parameters. It is this sum that must be compared against external geophysical constraints to assess physical plausibility.

11.3.1 Parameter Elasticities

The closed-form expression (Equation 11.3) reveals that $\dot{M}_0$ is highly sensitive to all three GR parameters, but with markedly different elasticities:

Table 11.2: Elasticity of moment release rate to GR parameters.

Parameter	Elasticity	Interpretation
$M_{\max}$	$\varepsilon \approx 8$–10	1% increase → 8–10% increase in $\dot{M}_0$
b-value	$\varepsilon \approx -3$ to -4	1% increase → 3–4% decrease in $\dot{M}_0$
λ_n	$\varepsilon = 1.0$	Linear (1:1) relationship

When multiple zones have uncertain parameters, combinations of high λ_n × low b × high $M_{\max}$ give very high $\dot{M}_0$, while the opposite gives very low values. Taking extreme combinations simultaneously across all zones compounds the problem—the implied range of total moment release can span orders of magnitude.

11.4 Sensitivity to Maximum Magnitude Discretization

The extreme sensitivity of $\dot{M}_0$ to $M_{\max}$ — with elasticities of 8–10 (Table 11.2) — means that the discretization of the $M_{\max}$ distribution in the logic tree has a direct effect on the computed moment release. Because the moment-magnitude relation is exponential, the moment release function $\dot{M}_0(M_{\max})$ is strongly convex, and the way $M_{\max}$ uncertainty is represented in the logic tree therefore matters.

Technical Detail: Jensen's Inequality and Discretization Bias

Because $\dot{M}_0$ is convex in $M_{\max}$, Jensen's inequality gives:

$$E\left[\dot{M}_0(M_{\max})\right] \geq \dot{M}_0(E[M_{\max}]) \tag{11.5}$$

Using arithmetic mean $M_{\max}$ underestimates expected moment release.

For uniform distribution spanning 1 magnitude unit:

- Three-point discretization: underestimates by 5.6%
- Five-point discretization: underestimates by 2.3%
- Continuous integration: no bias

11.5 Implications for Logic Tree Construction

The preceding results demonstrate that the NSHM 2020 source model produces a physically plausible moment release at its central parameter values, but the strong sensitivity to GR parameters — particularly $M_{\max}$ and b — means that many parameter combinations in the logic tree will imply moment release rates far outside the geodetically constrained range. This section considers how the moment constraint can be used operationally: first as a basis for identifying and downweighting implausible branches, and second as a mechanism for inducing physically motivated correlations between source zones.

11.5.1 Pruning Invalid Branches

For a logic tree with N source zones, each characterised by K branches over the three GR parameters $(\lambda_n, b, M_{\max})$, the full end-branch space contains K^{3N} combinations. For the UK NSHM with 22 zones and even a modest $3 \times 3 \times 4$ discretization per zone, this is astronomically large. Many of these combinations will imply total moment release rates that are orders of magnitude above or below the Main et al. (1999) constraint.

Figure 11.2 illustrates this for the UK NSHM 2020 model. Panel (a) shows the distribution of $\log_{10} \dot{M}_{\text{total}}$ across the full logic tree: the 5th–95th percentile range spans nearly a full order of magnitude, with the tails extending well beyond the geodetic constraint. Panel (b) decomposes the distribution by $M_{\max}$ value, confirming the dominant role of maximum magnitude in driving the spread. Panels (c) and (d) compare the implied moment release against external constraints, showing that the central estimate is consistent with observations but the tails include physically implausible combinations.

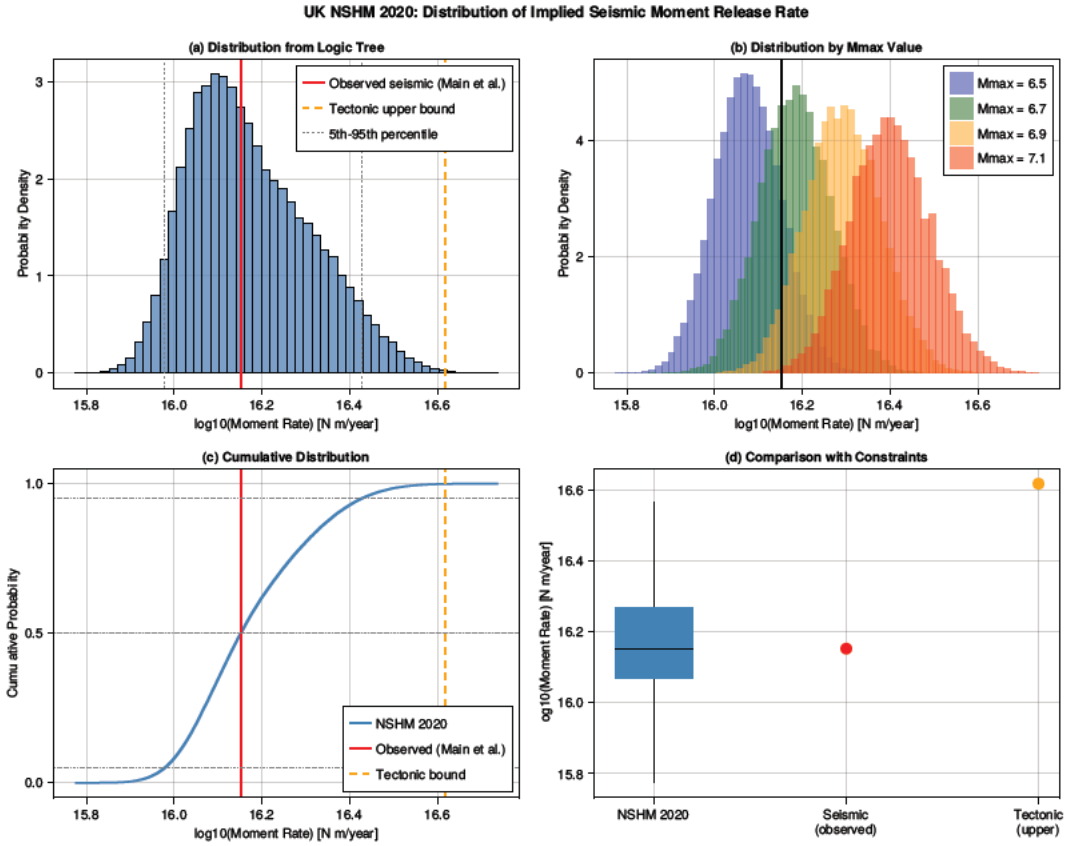

Figure 11.2: Distribution of implied seismic moment release rate across the UK NSHM 2020 logic tree. (a) Histogram of $\log_{10} \dot{M}_{\text{total}}$ from Monte Carlo sampling of all zone parameters. The observed seismic moment rate from (Main et al., 1999) and the tectonic upper bound are shown as vertical lines. (b) Decomposition by $M_{\max}$ value, showing the dominant effect of maximum magnitude on total moment release. (c) Cumulative distribution function. (d) Box plot comparison with external constraints.

Rather than discarding implausible branches outright — which would require choosing hard cut-off thresholds — it is more natural to apply the constraint as a soft penalty via importance sampling. Figure 11.3 illustrates the effect: panel (a) shows how the weight retention factor varies with the deviation of total moment release from the target, for different constraint strengths σ. Tighter constraints (smaller σ) reject more aggressively. Panel (d) shows that even a moderate constraint ($\sigma = 0.3$) reduces the effective number of logic tree combinations by approximately an order of magnitude, concentrating the hazard calculation on physically plausible parameter sets.

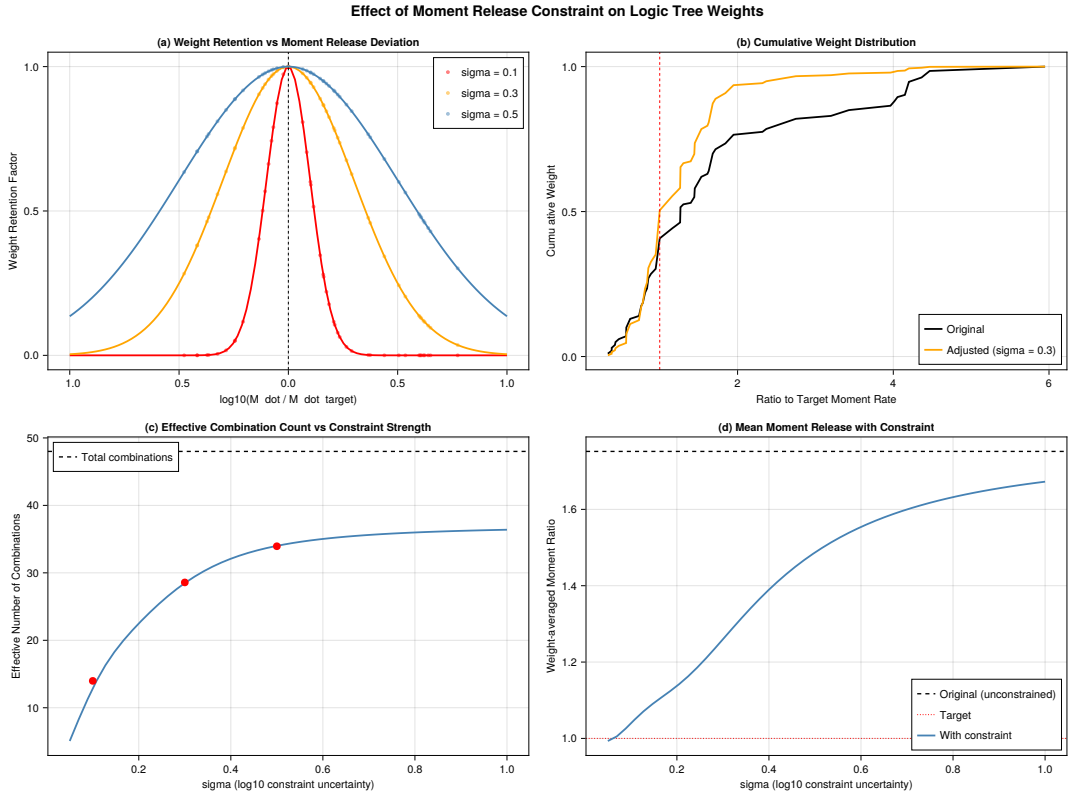

Figure 11.3: Effect of the moment release constraint on logic tree branch weights. (a) Weight retention factor as a function of the ratio $\dot{M}_{\text{total}}/\dot{M}_{\text{target}}$, for different constraint strengths σ. (b) Cumulative weight distribution before and after applying a moderate constraint ($\sigma = 0.3$). (c) Effective number of retained combinations as a function of constraint strength. (d) Mean moment release of the constrained ensemble as a function of σ, showing convergence toward the target.

11.5.2 Post-hoc Importance Sampling: Recommended Approach

The procedure has five steps:

1. **Independent zone calibration.** Run independent inversions for each zone to obtain marginal posteriors $p(\lambda_{n,z}, b_z \mid \text{data}_z)$ using standard methods (PMLM or Bayesian). Define $M_{\max}$ distributions from geological/tectonic considerations.
2. **Specify the moment constraint.** Express the constraint as a target distribution on the total moment release, for example:

$$\dot{M}_{\text{total}} \sim \text{LogNormal}\left(\mu = \ln\left(\dot{M}_{\text{target}}\right), \sigma = \sigma_{\ln \dot{M}}\right) \tag{11.6}$$

   where $\sigma_{\ln \dot{M}} \approx 0.3$–$0.5$ reflects uncertainty in the geodetic constraint. The Gaussian form means the constraint is applied as a soft penalty rather than a hard cut-off.
3. **Generate proposal samples and compute importance weights.** Draw K independent realisations of the complete source model from the unconstrained posteriors. For each realisation $k = 1, \ldots, K$, draw $(\lambda_{n,z}, b_z)$ from each zone's marginal posterior and $M_{\max,z}$ from its $M_{\max}$ distribution, compute the implied total moment release $\dot{M}^{(k)}_{\text{total}} = \sum_z \dot{M}^{(k)}_{0,z}$ using Equation 11.3, and assign an importance weight:

$$w_k = \exp\left(-\frac{1}{2}\left(\frac{\ln \dot{M}_{\text{total}}^{(k)} - \ln \dot{M}_{\text{target}}}{\sigma_{\ln \dot{M}}}\right)^2\right) \tag{11.7}$$

Realisations whose total moment release is close to the target receive weights near 1; those far from the target are strongly downweighted. After computing all K weights, normalise so that $\sum_k \tilde{w}_k = 1$.

4. **Resample.** The K proposal samples now have unequal normalised weights $\tilde{w}_k$. To obtain a set of equally-weighted samples from the moment-constrained distribution, apply Sampling Importance Resampling (SIR): draw N indices from $\{1, ..., K\}$ *with replacement*, where the probability of selecting index k is $\tilde{w}_k$. Samples with high weights (those consistent with the moment constraint) are selected multiple times; samples with low weights are rarely or never selected. The resulting N resampled realisations each carry equal weight $1/N$ and are approximately distributed according to the moment-constrained posterior. Each realisation is a complete, jointly drawn set of $(\lambda_{n,z}, b_z, M_{\max,z})$ across *all* zones, so cross-zone correlations induced by the constraint are preserved in the resampled ensemble. The effective sample size $\text{ESS} = 1/\left(\sum_k \tilde{w}_k^2\right)$ should be monitored to ensure that the proposal distribution has sufficient overlap with the constrained target; low ESS indicates that too few proposals are consistent with the constraint and K should be increased.
5. **Use in hazard calculations.** Each resampled realisation defines one complete source model that can be passed directly to the hazard engine. The mean hazard is computed by averaging over realisations, and fractile hazard is obtained from the empirical distribution of realisation-level hazard values.

The importance sampling approach is attractive because it preserves the marginal uncertainty on each zone's parameters — no zone is forced to adopt a particular rate or b-value — while inducing physically motivated correlations between zones. Realisations in which one zone contributes disproportionately to the total moment budget are only retained if other zones compensate with lower contributions, so the resampled ensemble naturally exhibits the negative inter-zone correlations discussed in Section 10.4. The approach is also operationally simple: it requires only a post-processing step on the existing unconstrained logic tree, with no modification to the hazard engine itself.

11.5.3 Hierarchical Models vs. Post-hoc Constraints

Two distinct strategies exist for incorporating the moment constraint into the PSHA workflow (Table 11.3). They differ fundamentally in *when* the constraint is applied: after zone calibration (post-hoc) or during it (hierarchical).

Approach 1: Post-hoc importance sampling

Each source zone is calibrated independently using existing methods (PMLM or Bayesian), producing marginal posteriors for (λ_n, b) per zone. The logic tree is constructed as normal, with each zone contributing its own set of parameter branches and weights. The moment constraint is then applied as a post-processing step: end-branch combinations across all zones are sampled, the implied total moment release $\dot{M}_{\text{total}}$ is computed for each combination, and combinations are re-weighted according to their consistency with geodetic constraints. Combinations that imply physically implausible total moment release are downweighted; plausible combinations retain their weight. The correlations between zones emerge from this re-weighting—if a sample assigns a high rate to one zone (which dominates moment release), it is only retained if other zones have correspondingly lower rates.

The re-weighted branches remain in the standard format expected by hazard software: a set of $(\lambda, b, M_{\max})$ values per zone with an associated weight. The hazard engine does not require modification—it processes each end-branch combination exactly as it would for an unconstrained logic tree. The only change is a pre-processing step that adjusts the branch weights before they are passed to the hazard calculation. This approach can also be applied retrospectively to existing studies by re-weighting their logic tree branches.

The main practical consideration is combinatorial: with K branches per parameter across N zones, the full end-branch space is too large to enumerate. The implementation therefore uses Monte Carlo sampling from the marginal posteriors with importance re-weighting and resampling, as described above. The number of resampled realisations N can be chosen to balance computational cost against sampling precision, and need not be large — a few hundred realisations typically suffice for stable estimates of mean hazard and key fractiles.

Approach 2: Joint hierarchical inversion

All zones are calibrated simultaneously within a single hierarchical model. Population-level hyperparameters (μ_β, σ_β, $\mu_{\ln\lambda}$, $\sigma_{\ln\lambda}$, ρ) govern the distribution from which zone-level parameters are drawn, and moment release constraints can be incorporated directly into the joint prior. Each posterior draw provides a complete, internally-consistent set of (λ_z, b_z) values across all zones, with correlations arising naturally from the shared hyperparameters and physical constraints.

This approach has fundamental advantages over post-hoc re-weighting: all posterior samples are physically plausible by construction (no wasted samples), the correlation structure is estimated from the data rather than imposed after the fact, and data-poor zones benefit from partial pooling with data-rich zones. However, integrating hierarchical posterior samples into existing PSHA workflows may require some adaptation. Standard hazard software expects a discrete logic tree with enumerated branches and weights, whereas the hierarchical model produces a cloud of correlated posterior samples. If the software supports a Monte Carlo mode—where each posterior draw defines one complete

source model realisation and the hazard engine is run once per draw—then integration is straightforward. If the software strictly requires a discrete logic tree, the posterior samples must first be translated into that format, for example by clustering samples into representative parameter combinations and assigning weights proportional to cluster membership. The critical requirement in either case is that each hazard realisation uses a *joint* draw of parameters across all zones; decomposing the posterior back into independent marginals per zone would discard the correlations that the hierarchical model was designed to capture.

Table 11.3: Comparison of practical implementation considerations for the two approaches to incorporating cross-zone correlations in PSHA logic trees.

Aspect	**Post-hoc importance sampling**	**Hierarchical model**
Zone calibration	Independent (existing workflow)	Joint (new framework)
Correlations introduced	At branch re-weighting stage	At calibration stage
Hazard engine changes	None—standard branch input	None if sample-based input supported; translation needed for discrete logic tree formats
Main new requirement	Pre-processing script	Bayesian hierarchical sampler
Retrofit to existing studies?	Yes	No—requires re-calibration

A practical consideration for both approaches is that the moment release constraint must reflect the spatial domain of the analysis. The Main et al. (1999) estimate of $\dot{M}_{\text{seismic}} = 1.42 \times 10^{16}$ N·m/year (Table 11.1) applies to the UK as a whole, but site-specific hazard assessments — such as those for nuclear installations — typically model only the sources within a radius of approximately 300 km from the site. The moment release target for such a study must be scaled accordingly, for example by summing the zone-level contributions from only those zones that fall within (or partially overlap) the analysis domain. This is straightforward in principle — the zone-level moment rates are already computed as part of the importance sampling procedure — but care is needed to ensure that the target is consistent with the source model being constrained.

Moment Constraint Summary:

- Current UK practice overstates epistemic uncertainty by treating zone parameters as independent
- The moment constraint induces **negative** correlations between zones (high in one → low in others)
- Post-hoc importance sampling should not pose significant challenges to implementation within existing hazard software, and can be applied retrospectively to existing studies
- Hierarchical modelling has fundamental advantages in statistical coherence and efficiency, but may require creative adaptation for integration with existing PSHA tools

12 Theme C: Joint Inversion

12.1 Motivation

The analyses presented in earlier sections of this report provide strong empirical motivation for a hierarchical approach to seismicity parameter estimation. The zone distinguishability tests in Section 10 demonstrated that adjacent UK source zones have statistically indistinguishable b-values in the vast majority of cases — 0% of zone pairs under PMLM fitting, and only 16% under unconstrained MLE — while activity rates differ significantly for 65–68% of adjacent pairs. This asymmetry suggests that the b-value is a regional parameter that should be shared (or at least partially pooled) across zones, whereas activity rates capture genuine spatial variation that should be preserved.

The standard approach in UK practice is to calibrate each source zone independently: zone z receives its own data set, and the Gutenberg–Richter parameters (λ_z, b_z) are estimated without reference to any other zone. This approach is statistically inefficient. Many UK zones contain fewer than 20 events above the completeness threshold, and some have fewer than 10. With so few data, zone-specific b-value estimates have large uncertainties and are strongly influenced by one or two outlier events. The zone distinguishability results confirm that most of this apparent between-zone b-value variation is noise rather than signal.

Hierarchical (multi-level) Bayesian models offer a principled solution to this problem through *partial pooling*. Rather than treating zone parameters as either fully independent (no pooling) or fully shared (complete pooling), a hierarchical model treats zone-specific parameters as draws from a common population distribution. This provides three key benefits:

1. **Shrinkage**: Zone estimates are pulled toward the population mean, with the strength of the pull inversely proportional to the zone's sample size. Data-sparse zones borrow strength from data-rich zones, while well-constrained zones are left largely unchanged.
2. **Between-zone variability estimation**: The population-level standard deviation (σ_b) is estimated from the data, providing a direct measure of how much genuine b-value variation exists across zones — answering the question that the zone distinguishability tests addressed indirectly.
3. **Correlation structure**: Population-level correlations between parameters (e.g., between $\ln \lambda$ and β) can be estimated, informing the correlation structure needed for logic tree discretization (Section 9).

The connection to the moment release analysis in Section 11 is also important. When zone parameters are calibrated independently, the implied total seismic moment release rate

across all zones is unconstrained and may be physically implausible. A hierarchical model can incorporate external constraints — such as geodetic moment rate estimates — directly within the calibration, rather than as a post-hoc correction.

> **Historical Note:** The idea of treating zone b-values as draws from a common distribution is not new. Dyck (1985) (Section 4.7.3.2) anticipated hierarchical seismicity modelling, treating zone b-values as "realizations of a random variable with normal distribution $\mathcal{N}(m_B, \sigma_B^2)$." He called this approach "empirical Bayes, because the prior distribution of each b_x is determined empirically" from the catalogue data. What has changed since 1985 is the availability of computational tools — particularly Markov chain Monte Carlo (MCMC) — that make full Bayesian hierarchical inference practical for models of this complexity.

12.2 Hierarchical Model Structure

The hierarchical model is specified at three levels: a population (hyper) level that describes the distribution of parameters across zones, a zone level where each zone's parameters are drawn from the population distribution, and a data level where observed earthquake counts follow a Poisson process conditioned on the zone parameters.

12.2.1 Population level

The population-level hyperparameters describe the joint distribution of zone parameters:

$$\mu_{\ln\lambda}, \quad \sigma_{\ln\lambda}, \quad \mu_\beta, \quad \sigma_\beta, \quad \rho \tag{12.1}$$

where $\mu_{\ln\lambda}$ and $\sigma_{\ln\lambda}$ are the mean and standard deviation of the log activity rates across zones, μ_β and σ_β are the corresponding parameters for the GR slope β, and ρ is the correlation between $\ln\lambda_z$ and β_z at the population level. Priors on these hyperparameters are:

$$\begin{aligned}
\mu_\beta &\sim \mathcal{N}(\ln 10, (0.1\ln 10)^2) && [\text{centered on } b = 1.0] \\
\sigma_\beta &\sim \text{HalfNormal}(0.15\ln 10) && [\text{weakly informative}] \\
\mu_{\ln\lambda} &\sim \mathcal{N}(0, 2^2) && [\text{vague}] \\
\sigma_{\ln\lambda} &\sim \text{HalfNormal}(1.0) && [\text{vague}] \\
\rho &\sim \text{Uniform}(-0.9, 0.9)
\end{aligned} \tag{12.2}$$

The prior on μ_β is centred on $b = 1.0$ with a standard deviation of 0.1 in b-value units (≈ 0.23 in β units), which is broadly consistent with the Johnston prior used in PMLM but expressed at the population rather than zone level. The prior on σ_β allows the data to determine how much genuine between-zone b-value variability exists.

12.2.2 Zone level

Zone-specific parameters are drawn from the population distribution. In the most general (correlated) variant:

$$\ln \lambda_z \sim \mathcal{N}(\mu_{\ln\lambda}, \sigma^2_{\ln\lambda}) \tag{12.3}$$

$$\beta_z \mid \ln \lambda_z \sim \mathcal{N}\left(\mu_\beta + \rho\frac{\sigma_\beta}{\sigma_{\ln\lambda}}(\ln\lambda_z - \mu_{\ln\lambda}),\quad \sigma^2_\beta(1-\rho^2)\right) \tag{12.4}$$

This conditional specification induces a bivariate normal population distribution for $(\ln\lambda_z, \beta_z)$ with correlation ρ. When $\rho = 0$, the β_z draws are independent of $\ln\lambda_z$ and the model reduces to the uncorrelated hierarchical variant.

12.2.3 Data level

Given zone parameters, the observed counts in magnitude bin i for zone z follow the Poisson process:

$$n_{i,z} \sim \text{Poisson}\big(\lambda_{m,z} \cdot t_{i,z} \cdot A_z \cdot \Delta F_{i,z}\big) \tag{12.5}$$

where $\lambda_{m,z}$ is the activity rate density for zone z (events per unit area per year above $M_{\min}$), $t_{i,z}$ is the effective observation period for magnitude bin i in zone z (accounting for magnitude-dependent completeness), A_z is the zone area, and $\Delta F_{i,z}$ is the probability mass in magnitude bin i from the truncated Gutenberg–Richter distribution with slope β_z.

12.2.4 Non-centred parameterisation

For efficient MCMC sampling, the zone-level β parameters use a non-centred parameterisation:

$$\beta_z = \mu_\beta + \sigma_\beta \cdot \beta_{\text{raw},z}, \qquad \beta_{\text{raw},z} \sim \mathcal{N}(0,1) \tag{12.6}$$

This avoids the "funnel" geometry that arises when σ_β is small and the zone-level parameters are tightly clustered around the population mean. The non-centred form ensures that the sampler explores the posterior efficiently even when the data are consistent with a near-zero σ_β (i.e., when all zones share essentially the same b-value).

12.3 Model Variants

To understand the contribution of each model component, four variants of increasing complexity are implemented:

1. **Independent**: Each zone is fitted with its own $(\ln\lambda_z, \beta_z)$ pair, using vague priors. This is the standard zone-by-zone approach used in current practice and serves as the baseline.
2. **Shared b**: A single β_{shared} is used across all zones, while activity rates λ_z remain zone-specific. This is the most restrictive pooling model: it assumes zero between-zone b-value

variability ($\sigma_\beta = 0$) and is appropriate if the zone distinguishability tests are taken at face value.

3. **Hierarchical**: Zone-specific β_z values are drawn from a common $\mathcal{N}\left(\mu_\beta, \sigma_\beta^2\right)$ distribution with σ_β estimated from the data. This provides partial pooling: the degree of shrinkage is determined automatically by the data, lying somewhere between the independent and shared-b extremes.
4. **Correlated hierarchical**: As above, but with a bivariate population distribution for $(\ln \lambda_z, \beta_z)$ with correlation ρ estimated from the data (Equation 12.3, Equation 12.4). This is the most general variant and can recover population-level correlations between activity rate and b-value.

A fifth variant, the **constrained** model, extends the hierarchical structure by adding a log-normal penalty on the total seismic moment release rate, connecting to the moment constraint framework described in Section 11. This incorporates the external geodetic constraint directly within the calibration rather than as a post-hoc correction.

12.4 Software Implementation

The hierarchical models are implemented in two independent software frameworks to provide cross-validation:

Julia (Turing.jl). The primary implementation uses Turing.jl, a probabilistic programming framework for Julia. Each model variant is defined as a Turing `@model` function that accepts a vector of `ZoneData` structures (containing binned magnitude counts, observation periods, and zone areas) and returns the sampled parameters. Inference is performed using the No-U-Turn Sampler (NUTS) with 4 chains of 2000 samples each. The Julia implementation is consistent with the rest of the analysis pipeline developed for this project and directly reuses the data structures and completeness models from the single-zone analyses.

Stan. A complementary implementation in Stan provides an independent check on the Turing.jl results. The Stan model (`gr_hierarchical_zones.stan`) uses the non-centred parameterisation (Equation 12.6) and additionally implements the Full Bayesian approach from Section 6, with latent true magnitudes for each event and an observation model accounting for Grünthal ML-to-Mw conversion uncertainties. Stan functions are defined for the forward and inverse Grünthal conversion, the conversion slope (Jacobian), and the total uncertainty computation. Inference uses Stan's adaptive HMC (NUTS) sampler. The Stan implementation is called from R via `rstan`.

The key difference between the two implementations is that the Turing.jl version operates on pre-binned magnitude counts (matching the Weichert/PMLM framework), while the Stan version operates on individual event magnitudes with latent variables (matching the Full

Bayesian framework). Agreement between the two provides confidence that the hierarchical structure is correctly implemented and that the results are not sensitive to the choice of data representation.

12.5 Synthetic Validation

Before applying the hierarchical model to the UK catalogue, a synthetic validation exercise is conducted to verify that the model can correctly recover known parameters. Synthetic data are generated for 8 zones with the following population parameters:

- Population mean b-value: $\mu_b = 1.0$
- Between-zone b-value standard deviation: $\sigma_b = 0.12$
- Population correlation: $\rho(\ln \lambda, \beta) = 0.4$
- Observation period: $T = 50$ years per zone
- Magnitude range: M_w 3.0 to 6.5

Zone-specific parameters are drawn from the correlated bivariate population distribution, and event catalogues are generated from the resulting inhomogeneous Poisson processes. The synthetic zone sample sizes range from data-sparse (< 20 events) to data-rich (> 100 events), reflecting the range observed in the UK source model.

All four model variants are fitted to the same synthetic data set. Figure 12.1 shows the resulting parameter estimates compared to the known true values.

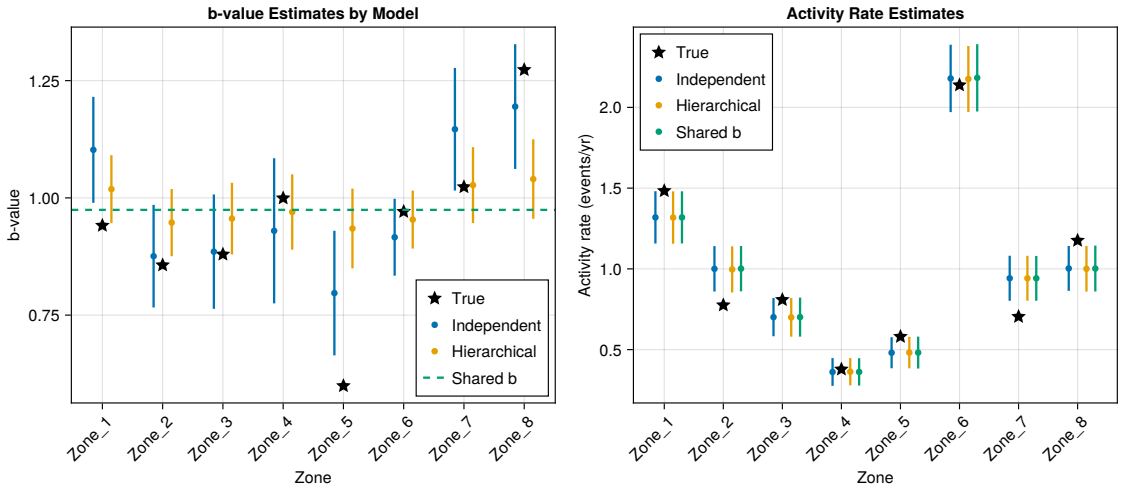

Figure 12.1: Comparison of b-value (left) and activity rate λ_n (right) estimates across three model variants — independent (no pooling), hierarchical (partial pooling), and shared b (complete pooling) — fitted to the same synthetic data set with 8 zones. True parameter values are shown as black stars. For b-values, the hierarchical model (orange) produces reduced posterior uncertainty relative to the independent model (blue) through shrinkage toward the population mean, though neither model is uniformly closer to the true values across all zones. Activity rate estimates are similar across all three models, as expected since the rate is primarily constrained by the observed event counts in each zone regardless of how b is treated. The shared-b model (green dashed line in left panel; green points in right panel) imposes a single b-value across all zones. Error bars show ± 1 posterior standard deviation.

The independent model recovers the true parameters for data-rich zones but produces noisy, uncertain estimates for data-sparse zones. The shared-b model provides a precise but potentially biased estimate — it forces all zones to share a single b-value, which is correct on average but cannot capture genuine between-zone differences. The hierarchical model provides the best compromise: its estimates are pulled toward the population mean (reducing noise) while still allowing zone-specific departures where the data support them.

The shrinkage behaviour is illustrated in Figure 12.2, which plots each zone's hierarchical b-value estimate against its independent estimate. Under the hierarchical model, all estimates are pulled toward the population mean (dotted lines), with the strength of the pull depending on each zone's sample size. Zones with few events (small points) experience strong shrinkage, while data-rich zones (large points) remain close to the 1:1 line.

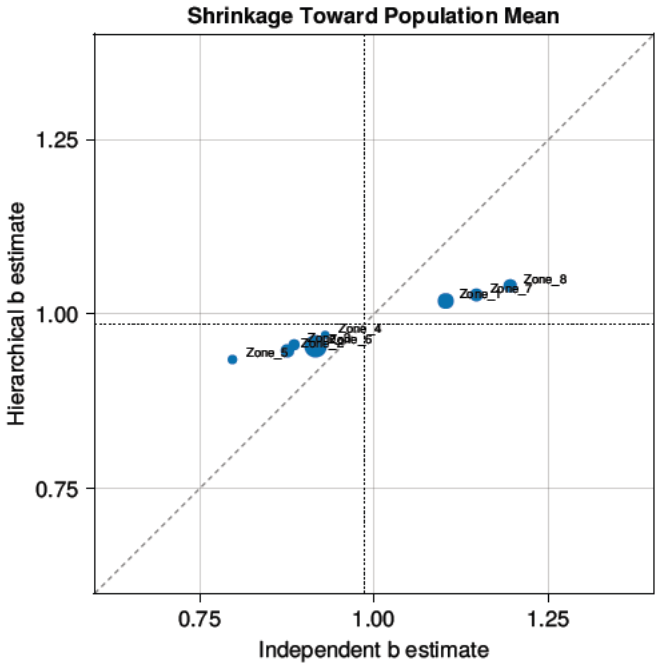

Figure 12.2: Shrinkage effect in the hierarchical model. Each point represents one zone, plotted as its independent b-value estimate (horizontal axis) versus its hierarchical estimate (vertical axis). Point sizes are proportional to the number of events in each zone. The dashed diagonal line is the 1:1 reference; dotted lines mark the estimated population mean μ_b. Data-sparse zones (small points) are pulled strongly toward the population mean, while data-rich zones (large points) remain close to their independent estimates.

This shrinkage is the central mechanism of the hierarchical model. The effective shrinkage weight for zone z can be approximated as:

$$\hat{\theta}_z^{\text{hier}} \approx w_z \hat{\theta}_z^{\text{indep}} + (1 - w_z)\hat{\theta}^{\text{pop}}, \qquad w_z = \frac{\sigma_\beta^2}{\sigma_\beta^2 + \sigma_z^2} \tag{12.7}$$

where σ_z^2 is the sampling variance of the zone-specific estimate (inversely related to sample size) and σ_β^2 is the between-zone variance. When $\sigma_z^2 \gg \sigma_\beta^2$ (data-sparse zone), $w_z \approx 0$ and

the estimate shrinks to the population mean. When $\sigma_z^2 \ll \sigma_\beta^2$ (data-rich zone), $w_z \approx 1$ and the zone retains its own estimate.

For the synthetic example, the hierarchical model recovers the population hyperparameters well: $\hat{\mu}_b \approx 1.0$ and $\hat{\sigma}_b \approx 0.12$, consistent with the generating values. The correlated variant also recovers the population correlation $\hat{\rho} \approx 0.4$.

12.6 UK Application

The hierarchical framework validated above is now applied to the actual UK earthquake catalogue. The BGS catalogue is loaded with Mw conversion using the Grünthal et al. (2009) relation and filtered for completeness using the same model as the zone distinguishability tests (Section 10.3). Events are assigned to the 22 UK NSHM source zones by point-in-polygon test. All qualifying zones are included in the hierarchical pool, with the bipartite zones MMCW and MENA treated as single Gutenberg–Richter distributions as a simplification (see the discussion of bipartite zones below for caveats). This preserves the total moment budget across all zones, which is essential for the moment constraint comparison that follows. Zones with fewer than 3 events above the completeness threshold at $M_w \geq 3.0$ are excluded, as they provide negligible constraint on zone-specific parameters. The resulting analysis comprises approximately 18 zones spanning from Cornwall (CORN) to the Scottish Highlands (SC9), with zone sample sizes ranging from fewer than 5 events to more than 40.

Figure 12.3 shows the parameter estimates from the independent, hierarchical, and shared-b models applied to the UK data. The pattern is qualitatively similar to the synthetic validation, but with two important differences. First, the spread of independent b-value estimates is wider than in the synthetic case, reflecting the genuine data sparsity of individual UK zones. Second, the hierarchical shrinkage is correspondingly more pronounced: data-sparse zones are pulled substantially toward the population mean, while the few data-rich zones retain estimates closer to their independent values.

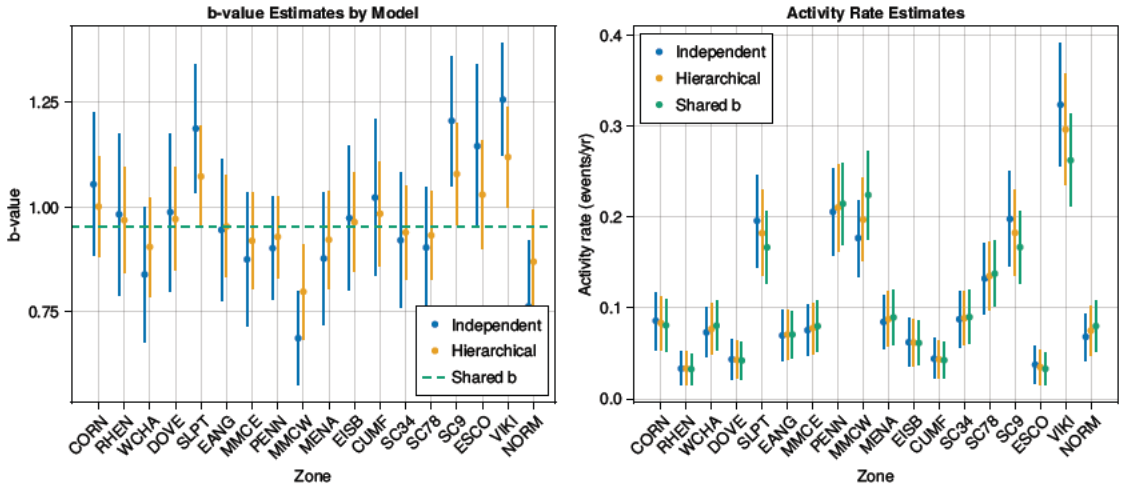

Figure 12.3: Comparison of b-value (left) and activity rate (right) estimates across three model variants applied to UK source zone data. The hierarchical model (orange) produces narrower uncertainty intervals than the independent model (blue) through shrinkage toward the population mean, while the shared-b model (green dashed line, left; green points, right) imposes a single b-value across all zones. Error bars show ± 1 posterior standard deviation.

The activity rate estimates (right panel) are largely insensitive to the model choice, consistent with the synthetic validation: the rate is primarily constrained by observed event counts, which are unaffected by how the b-value is treated. The shared-b model rates are indistinguishable from the hierarchical rates for most zones.

The shrinkage behaviour is illustrated more directly in Figure 12.4. Data-sparse zones (small points) experience strong shrinkage toward the population mean (dotted lines), while data-rich zones (large points) remain close to the 1:1 line. The overall pattern is consistent with the zone distinguishability tests in Section 10.3, which found that most zone b-values are statistically indistinguishable from the regional average.

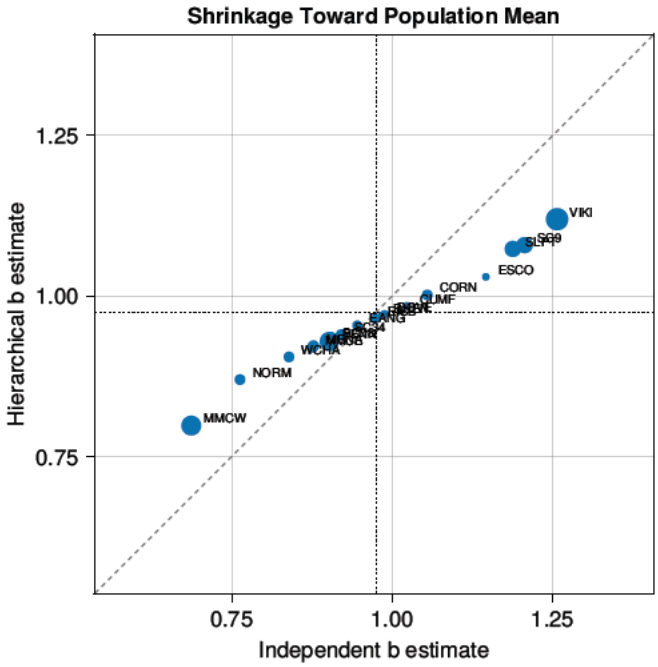

Figure 12.4: Shrinkage effect in the hierarchical model applied to UK source zones. Each point represents one zone, plotted as its independent b-value estimate versus its hierarchical estimate. Point sizes are proportional to zone sample size. Dotted lines mark the estimated population mean μ_b.

The population-level hyperparameters provide direct estimates of the quantities that UK practice has previously assumed. Figure 12.5 shows the posterior distributions for μ_b (the population mean b-value), σ_b (the genuine between-zone standard deviation), and ρ (the population-level correlation between activity rate and b-value).

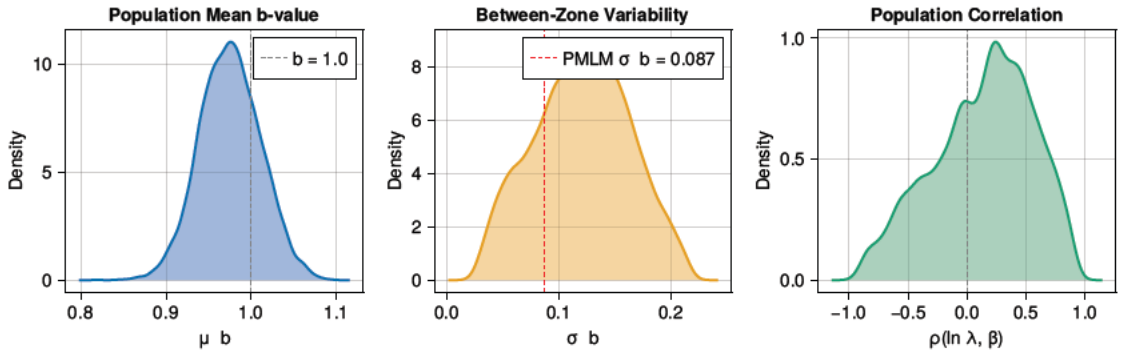

Figure 12.5: Posterior distributions for hierarchical model hyperparameters estimated from UK source zone data. Left: population mean b-value μ_b, with $b = 1.0$ reference (dashed). Centre: between-zone standard deviation σ_b, with the PMLM prior standard deviation $\sigma_b = 0.087$ (red dashed). Right: population correlation $\rho(\ln \lambda, \beta)$ between zone-level activity rate and b-value.

The estimated σ_b is particularly informative in the context of the PMLM prior discussion in Section 10.3. The UK PMLM applications use $\sigma_b = 0.087$ as a fixed value, assumed to be valid for the UK data. The hierarchical model instead learns σ_b directly from the data. If the posterior for σ_b is concentrated well below 0.087, this confirms that the PMLM prior is too

wide for UK zones — consistent with the zone distinguishability finding that most UK zones share a common b-value. Conversely, if σ_b is comparable to or exceeds 0.087, this would suggest that the PMLM prior is appropriately capturing genuine between-zone variability.

12.6.1 Inter-Zone Posterior Correlations

A critical advantage of the hierarchical framework over zone-by-zone analyses is that it provides the *complete joint posterior* over all zone parameters simultaneously, not merely independent marginals. The shared hyperparameters (μ_β, σ_β) induce posterior correlations between zone b-values: updating μ_β from the data shifts all zone estimates in the same direction, creating positive inter-zone correlations even without any explicit constraint.

Figure 12.6 shows the posterior correlation matrices for zone b-values and activity rates from the unconstrained correlated hierarchical model. The b-value correlations (left panel) are uniformly positive, reflecting the shared population mean — if the data suggest a higher μ_β, all zone β_z values shift upward together. The $\ln(\lambda)$ correlations (right panel) are weaker, because activity rates are primarily constrained by zone-specific event counts rather than the shared distribution.

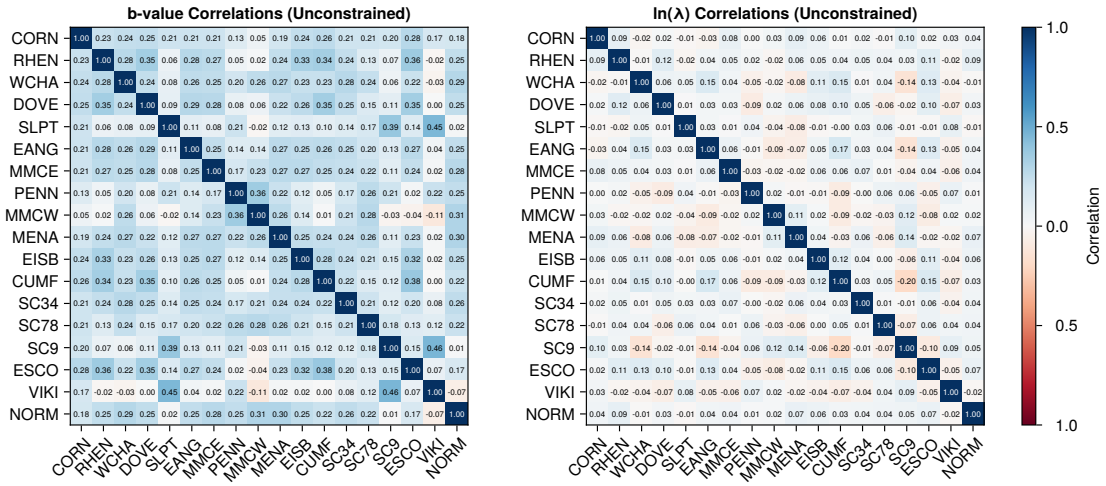

Figure 12.6: Inter-zone posterior correlations from the unconstrained correlated hierarchical model. Left: b-value correlations across zones. Right: $\ln(\lambda)$ correlations. The positive b-value correlations arise because all zones are linked through the shared population mean μ_β — updating μ_β shifts all zone estimates in the same direction. Colour scale: blue (negative) to red (positive).

This correlation structure becomes even richer when external constraints are applied. Adding a log-normal constraint on total seismic moment release ($\dot{M}_{\text{target}} = 1.42 \times 10^{16}$ N·m/yr, Main et al. (1999)) via importance sampling induces *additional* correlations: if one zone's implied moment rate is high (because its b-value is low or its rate is high), other zones must compensate to keep the total within the plausible range.

Figure 12.7 illustrates this effect. The left panel shows how the moment constraint shifts the implied total moment rate distribution toward the observational target. The centre and

right panels compare the inter-zone b-value correlation matrices before and after applying the constraint. The moment constraint strengthens the off-diagonal correlations, reflecting the physical trade-offs between zones that any zone-by-zone analysis necessarily misses.

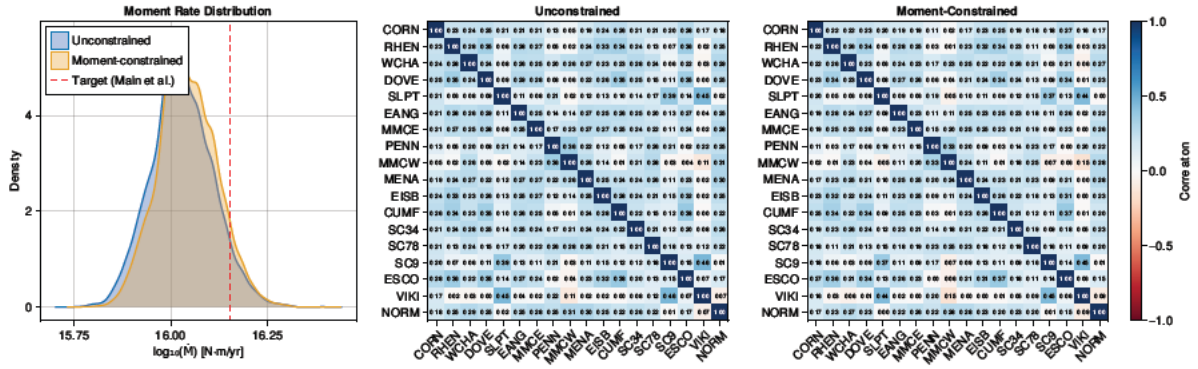

Figure 12.7: Effect of the seismic moment constraint on inter-zone correlations. Left: distribution of implied total seismic moment release rate before (blue) and after (orange) applying the constraint, with the observational target from Main et al. (1999) marked (red dashed). Centre: unconstrained b-value correlation matrix. Right: moment-constrained b-value correlation matrix. The constraint induces stronger inter-zone correlations reflecting the physical requirement that total moment release remain plausible.

This is what makes the hierarchical approach fundamentally more powerful than merely using a stronger prior on b within each zone independently. A zone-by-zone analysis — whether using Weichert MLE, PMLM, or even a strong Bayesian prior — produces independent posteriors that cannot capture trade-offs between zones. The hierarchical joint posterior preserves these trade-offs, and external constraints like moment release amplify them. For hazard calculations where the combined contribution from multiple sources matters, these inter-zone correlations directly affect the uncertainty in aggregate hazard estimates.

12.7 External Constraints

The hierarchical framework can be extended to incorporate external constraints on the model. The most natural constraint is on the total seismic moment release rate, connecting to the analysis in Section 11:

$$\ln \dot{M}_{\text{total}} \sim \mathcal{N}\left(\ln \dot{M}_{\text{target}}, \sigma^2_{\ln \dot{M}}\right) \tag{12.8}$$

where $\dot{M}_{\text{total}} = \sum_z \dot{M}_z$ is the total implied moment release across all zones (computed from the current zone parameters using Equation 11.3), $\dot{M}_{\text{target}}$ is the external constraint (e.g., from geodetic data, Table 11.1), and $\sigma_{\ln \dot{M}}$ controls the tightness of the constraint. This is implemented as an additional log-normal penalty term in the model, identical in form to the importance sampling weights used in the post-hoc approach but now incorporated directly into the posterior.

The advantage of incorporating the constraint within the calibration — rather than as a post-hoc re-weighting — is that the constraint influences the zone-level parameter estimates directly. Zones whose parameters imply an implausibly high moment release are pulled back during calibration, and the resulting posterior correlations between zones are induced naturally rather than imposed after the fact. The disadvantage is that the model must be re-fitted whenever the constraint changes, whereas the post-hoc approach can be applied to existing calibrations (Table 11.3).

The framework can also accommodate spatial smoothness constraints (requiring neighbouring zones to have similar parameters) and regional bounds on maximum magnitudes, although these extensions are not pursued here.

12.8 Treatment of Bipartite Zones

Two of the UK source zones — MMCW (Mid and Western Mainland Scotland) and MENA (Menai Strait) — use bipartite magnitude-frequency distributions in the NSHM 2020 model (Mosca et al., 2022). These consist of two Gutenberg–Richter distributions joined at a characteristic magnitude $M_c \approx 4.5$:

$$f_M(m) = \begin{cases} C_1 \beta_1 \frac{e^{-\beta_1 (m - M_{\min})}}{Z_1} & M_{\min} \leq m < M_c \\ C_2 \beta_2 \frac{e^{-\beta_2 (m - M_c)}}{Z_2} & M_c \leq m \leq M_{\max} \end{cases} \tag{12.9}$$

where Z_1 and Z_2 are normalisation constants and C_1, C_2 ensure continuity at M_c. This formulation allows for a steeper slope at lower magnitudes (larger β_1) and a shallower slope at higher magnitudes (smaller β_2), reflecting a change in the character of seismicity at the transition magnitude. Or, the model can have a single slope ($\beta_1 = \beta_2$) but a change in activity rate at M_c.

This poses a challenge for the hierarchical model, which assumes a single Gutenberg–Richter distribution for each zone. Fitting a single GR model to bipartite zone data yields biased estimates: the low b-value (0.55) reported for MMCW in Table 10.2 reflects the fitting procedure's attempt to accommodate both populations with a single slope, producing an estimate that is too shallow for the lower magnitudes and too steep for the upper magnitudes.

Due to the $10^{1.5m}$ scaling of seismic moment (Equation 11.2), the upper component of the bipartite distribution dominates the total moment release despite representing fewer events. This means that the moment constraint is particularly sensitive to the β_2 parameter, which is estimated from very few large events and therefore has substantial uncertainty.

For the hierarchical model, we recommend:

1. **Exclude bipartite zones from the hierarchical b-value pool.** Their b-values are not comparable with unimodal zones and would distort the population distribution.
2. **Use original bipartite parameters for moment release calculations.** The zone-level moment rate should be computed from the bipartite distribution using both components.
3. **Consider separate analysis above and below M_c.** If the lower segment ($m < M_c$) is of primary interest, it can be included in the hierarchical model using only events below M_c.
4. **Account for increased model uncertainty.** The bipartite formulation introduces additional parameters (β_1, β_2, M_c) and the associated epistemic uncertainty should be reflected in the logic tree.

Joint Inversion Summary:

- Hierarchical models provide a principled framework for partial pooling of b-values across UK source zones, directly addressing the finding that zone-specific b-values are statistically indistinguishable
- The hierarchical approach automatically calibrates the degree of shrinkage from the data: data-sparse zones borrow strength from the regional pool, while data-rich zones retain their own estimates
- The population-level standard deviation σ_b provides a direct, data-driven estimate of genuine between-zone b-value variability, calibrated from UK data rather than assumed from global compilations
- Application to UK data confirms that the between-zone b-value variability is small, consistent with the zone distinguishability tests, and provides a principled comparison point for the PMLM prior ($\sigma_b = 0.087$)
- The full joint posterior reveals inter-zone correlations induced by shared hyperparameters and moment constraints — information unavailable from independent zone analyses
- External constraints (e.g., total moment release) can be incorporated within the calibration, inducing physically motivated correlations between zones
- Bipartite zones (MMCW, MENA) require special treatment; when included with single-GR approximation, they preserve the total moment budget for constraint analysis

13 Theme C: Towards Data-Driven Source Zonation

The original scope for this research included investigation of flexible source boundaries, including data-driven approaches that use statistical algorithms to group earthquake locations into zones automatically, without relying on expert-drawn boundaries. While a full implementation of flexible-boundary inversion — where zone membership becomes a latent variable within the hierarchical model — is beyond the current scope, this section presents two complementary analyses that address the underlying question: *how sensitive are our results to the assumed zone boundaries?* The first approach uses a standard clustering algorithm to ask what zonation the data alone would suggest; the second works within the existing expert-defined boundaries but asks which of those boundaries are actually justified by the data.

13.1 Motivation: The Problem of Inherited Boundaries

Current UK source zone boundaries derive from geological and geophysical arguments: fault systems, tectonic domains, crustal thickness variations, Caledonian terrane boundaries, and similar considerations. While these arguments can *motivate* a boundary placement, the correlation – let alone causation – with observed seismicity patterns is rarely demonstrated quantitatively.

There is also significant anchoring bias in UK practice. Successive hazard models have made incremental adjustments to inherited zone geometries rather than questioning fundamental assumptions. The current NSHM 2020 zonation (Mosca et al., 2022) evolved from earlier models, which themselves built upon still earlier work. This creates a form of path dependence where current boundaries reflect accumulated expert judgment across decades, but the statistical support for these specific boundaries (as opposed to plausible alternatives) has not been formally evaluated.

This matters for epistemic uncertainty. Current logic trees include alternative b-values, activity rates, and maximum magnitudes, but typically treat zone boundaries as fixed. Yet the choice of boundaries may contribute as much epistemic uncertainty as parameter uncertainty within fixed zones. A purely data-driven approach provides a useful counterfactual: what would zone boundaries look like if derived solely from observed seismicity?

13.2 Approach 1: K-means Clustering as Proof of Concept

13.2.1 Method

K-means clustering is a standard partitioning algorithm that divides a set of observations into K groups by iteratively assigning each observation to the group whose centre (centroid) is nearest, then updating each centroid to be the average location of its assigned members. The algorithm repeats these two steps until the assignments stabilise. The result is K non-overlapping clusters, each defined by a centroid, where every observation belongs to the cluster whose centroid it is closest to. The key input is the number of clusters K, which must be specified in advance.

As a proof of concept, K-means clustering is applied to declustered epicentre locations from the UK catalogue. The number of clusters is set to $K = 22$, matching the current UK zone count, to enable direct comparison. To convert the set of cluster centroids into a complete spatial zonation, a Voronoi tessellation is constructed. A Voronoi tessellation partitions a plane around a set of seed points such that every location in the plane is assigned to the nearest seed point; the resulting regions (Voronoi cells) are convex polygons whose boundaries are equidistant between adjacent seed points. Here, the K-means centroids serve as the seed points, and the resulting Voronoi cells are clipped to the UK land and sea boundary to produce candidate source zones.

Important: This analysis is explicitly presented as a proof of concept with known limitations. The purpose is not to propose K-means zonation as superior to expert-defined zones, but to demonstrate that data-driven approaches yield different results —quantifying the degree to which current boundaries represent one choice among many compatible with the data.

13.2.2 Known Limitations

The following limitations are acknowledged:

- **Cluster shape assumption**: K-means assumes isotropic (spherical) clusters, but seismicity may follow linear features (faults, rift margins, topographic gradients) that are not well captured by spherical clusters
- **Equal event weighting**: All events contribute equally regardless of magnitude
- **Completeness variation**: No accounting for variable completeness (offshore regions have shorter complete periods)
- **Boundary geometry**: Voronoi cells have straight boundaries; real geological boundaries are irregular
- **Choice of K**: The number of clusters is arbitrary (though matched to current practice for comparison)

- **Two-dimensional**: No consideration of depth distribution
- **Stationarity**: Temporal stationarity assumed; cluster structure may evolve over geological time

More sophisticated clustering approaches would address some of these limitations. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) identifies clusters as regions of high point density separated by regions of low density, without requiring the number of clusters to be specified in advance and with the ability to identify arbitrarily shaped clusters — including the elongated patterns that seismicity often exhibits along fault zones. Gaussian mixture models replace the hard cluster boundaries of K-means with probabilistic assignments, fitting a mixture of Gaussian distributions to the data so that each event has a probability of belonging to each cluster rather than a binary assignment; this naturally accommodates clusters of different sizes and orientations. Spatial point process models go further by treating the earthquake locations as realisations of an underlying intensity function that varies continuously across space, avoiding the need for discrete zone boundaries altogether. These alternatives are noted as potential future extensions.

13.2.3 Results

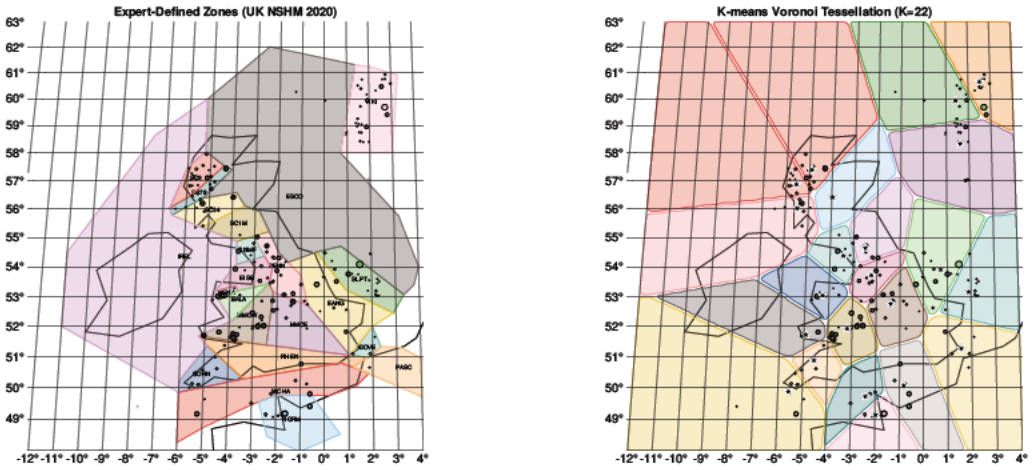

Figure 13.1: K-means clustering ($K = 22$) applied to UK seismicity (Mw ≥ 3.0, 206 events). Left: Voronoi cells from K-means centroids. Right: Current UK NSHM 2020 source zones. Events are coloured by cluster assignment (left) and expert zone (right). The data-driven boundaries bear limited resemblance to expert-defined zones.

Figure 13.1 presents the comparison between K-means-derived and expert-defined zones for the complete UK catalogue (Mw ≥ 3.0). The quantitative agreement rate – defined as the fraction of events where K-means cluster assignment is consistent with expert zone membership – is 75.6%.

This agreement rate reflects a specific methodology: since K-means clusters and expert zones do not share the same labels, a mapping between them must be established. For each

K-means cluster, the expert zone containing the majority of that cluster's events is identified as its "best match." The agreement rate is then the fraction of all events for which the K-means cluster assignment and the best-match expert zone assignment coincide. The result indicates that roughly one quarter of events would be assigned to different zones under the two approaches — events that fall on one side of an expert boundary but are grouped with events on the other side by the data-driven method.

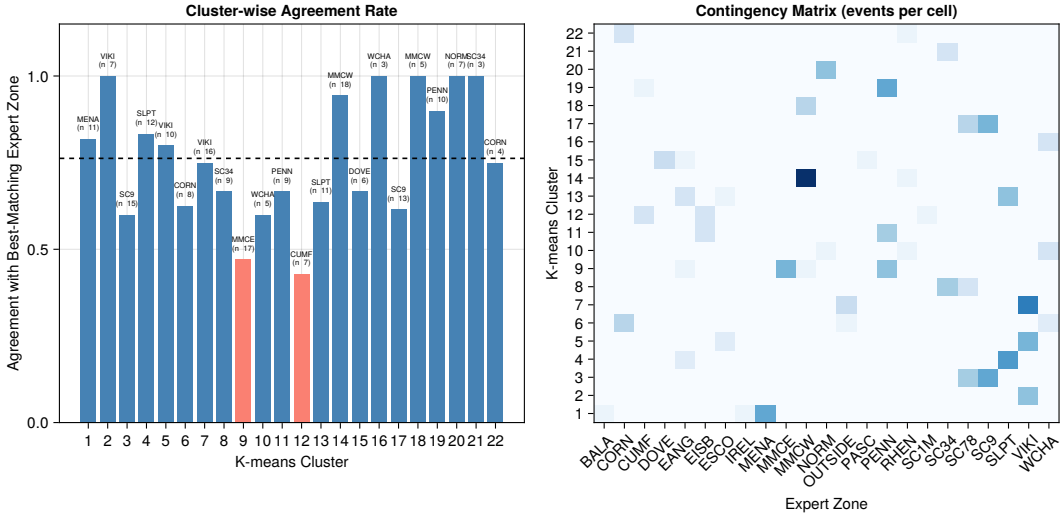

Figure 13.2: K-means agreement analysis showing cluster-by-cluster mapping between data-driven clusters and expert zones. Cluster assignments match expert zones for approximately 75% of events.

The analysis was repeated at multiple magnitude thresholds to assess stability:

Table 13.1: K-means vs expert zone agreement at different magnitude thresholds. Agreement decreases slightly with more data, suggesting the discrepancy is systematic rather than sampling noise.

Threshold	Events	Agreement Rate	Zones with Data
Mw ≥ 3.0	206	75.6%	18 / 22
Mw ≥ 2.5	373	73.1%	21 / 22
Mw ≥ 2.0	795	70.9%	21 / 22

13.2.4 Discussion

The declining agreement rate with increasing sample size (75.6% → 70.9%) is particularly informative. If the discrepancy were purely due to sampling variability, we would expect agreement to improve with more data. Instead, the persistent 25-30% disagreement indicates a systematic difference between data-driven and expert-derived zonation.

The comparison between data-driven and expert-defined zones is informative regardless of which approach is "better." Substantial differences would indicate that current boundaries encode significant expert interpretation beyond what the seismicity data alone would suggest. This interpretation may be entirely appropriate – but it represents epistemic uncertainty that is currently unquantified.

13.3 Approach 2: Merging Statistically Indistinguishable Zones

13.3.1 Motivation

Results from Section 10 demonstrate that 68% of UK source zones are statistically indistinguishable from the regional b-value. If zones cannot be distinguished based on the available data, maintaining them as separate entities:

- Increases parametric uncertainty (smaller sample sizes per zone)
- Creates an illusion of spatial resolution not supported by data
- Complicates logic tree construction with branches that the data cannot discriminate

A principled alternative is to merge zones that the data cannot distinguish, working within the existing boundary framework.

13.3.2 Method

The approach works by progressively simplifying the existing zonation, merging adjacent zones that the data cannot distinguish, in a process known as hierarchical agglomeration. The idea is straightforward: if two neighbouring zones have b-value estimates that are statistically indistinguishable — that is, if their posterior distributions overlap to such an extent that the data provide no evidence for a boundary between them — then the boundary is removed and the two zones are combined into one.

Zones are considered candidates for merging based on posterior overlap:

- Zones i and j are mergeable if their 80% credible intervals for b overlap substantially
- Alternatively, a Bayes factor comparing "same b-value" versus "different b-value" hypotheses can be computed
- Spatial adjacency is required: only neighbouring zones may be merged, preserving the spatial contiguity of the resulting zonation

The agglomeration proceeds as a greedy sequence of pairwise merges:

1. Start with the current 22 zones and estimate b-value posteriors for each
2. Compute a similarity score for every pair of adjacent zones, based on the overlap of their b-value posteriors
3. Merge the most similar pair, combining their earthquake catalogues into a single zone
4. Re-estimate b-value posteriors for all zones (since the merged zone now has a larger catalogue, its posterior will typically be narrower)
5. Repeat steps 2–4 until no remaining pair of adjacent zones exceeds the similarity threshold — that is, until all surviving zone boundaries are statistically justified by the data
6. Record the full merging sequence and corresponding similarity scores, which can be displayed as a dendrogram (a tree diagram showing which zones were merged and at what similarity level)

13.3.3 Results: b-value Similarity Only

As a proof of concept demonstrating how the hierarchical merging method works, we first apply it using b-value similarity alone. This illustrates the mechanics of the approach before introducing the more complete similarity measure that also accounts for activity rate differences.

The hierarchical merging analysis uses the Full Bayesian L5 Bayesian model (see Section 6) to estimate b-value posteriors for each zone, then iteratively merges the most similar adjacent pairs. Similarity is quantified using a normalised posterior overlap score ranging from 0 (clearly distinct) to 1 (statistically identical): a score of 1 means the two posteriors are effectively identical, while a score below a chosen threshold (here, approximately 0.6) indicates that the zones are sufficiently different to warrant separate treatment.

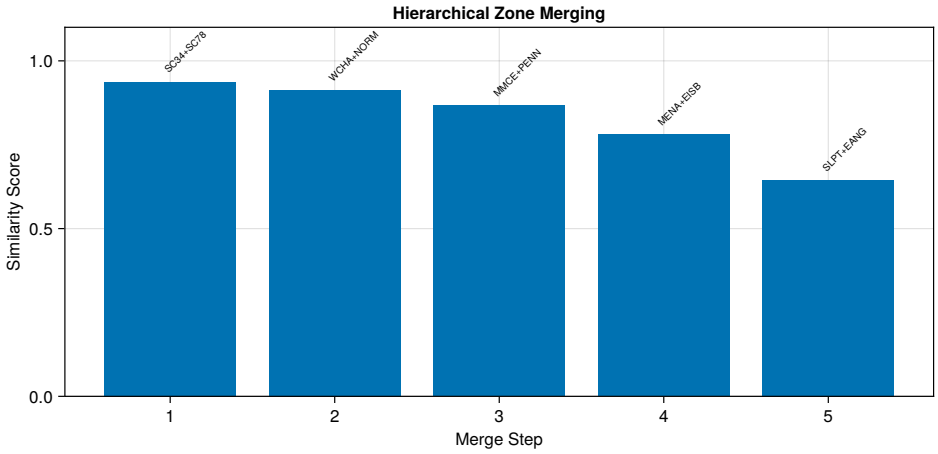

Figure 13.3: Similarity scores for hierarchical zone merging (Mw ≥ 3.0). Higher scores indicate greater statistical similarity. Five zone pairs exceed the similarity threshold and are merged.

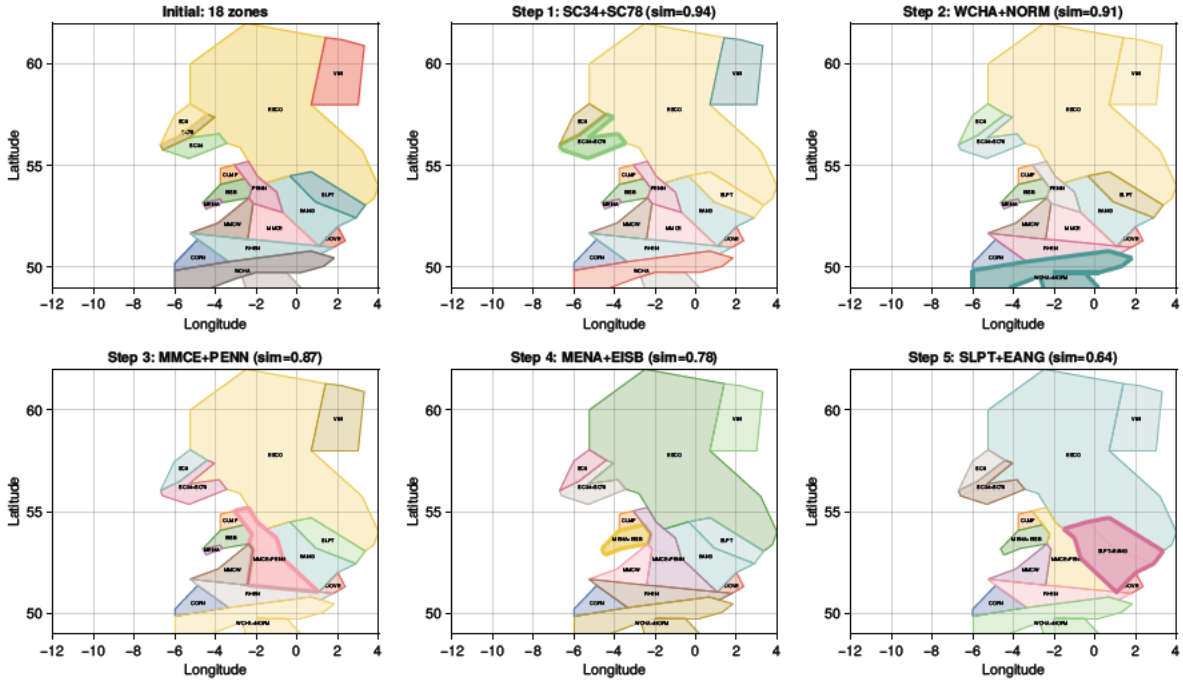

Figure 13.4: Evolution of zone configuration through the merging sequence. Panel 1 shows the initial 18 zones with sufficient data. Subsequent panels show the result of each merge, with the merged zone highlighted. The final configuration contains 13 zones.

Starting from the 22 UK NSHM 2020 zones, four zones have insufficient data for independent calibration at Mw ≥ 3.0: BALA, SC1M, IREL, and PASC. These require borrowed or regional parameters regardless of the methodology employed.

Of the 18 zones with sufficient data, the hierarchical merging identifies five pairs as statistically indistinguishable:

Table 13.2: Zone merging sequence for Mw ≥ 3.0 catalogue. All merges occur between adjacent zones with overlapping b-value posteriors.

Merge	Zone 1	Zone 2	Similarity	Combined b
1	SC34	SC78	0.94	0.86
2	WCHA	NORM	0.91	0.76
3	MMCE	PENN	0.87	0.80
4	MENA	EISB	0.78	0.89
5	SLPT	EANG	0.64	1.05

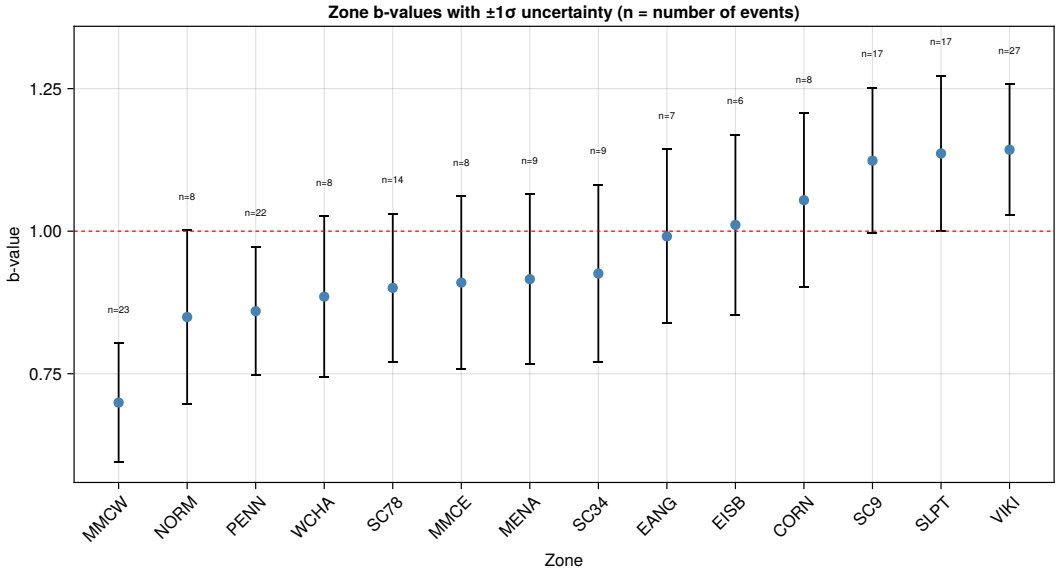

Figure 13.5: Zone b-values with 95% credible intervals from Full Bayesian L5 analysis (Mw ≥ 3.0). Zones with overlapping intervals are candidates for merging. The wide credible intervals reflect the limited data available per zone.

The sensitivity to magnitude threshold was assessed:

Table 13.3: Merging results at different magnitude thresholds. Lower thresholds provide more data per zone, enabling better discrimination between zones.

Threshold	Zones with Data	Merges	Final Zones
Mw ≥ 3.0	18	5	13
Mw ≥ 2.5	21	8	13
Mw ≥ 2.0	21	6	15

13.3.4 Results: Joint b-value and Rate Density Comparison

The b-value-only analysis above demonstrates how the merging method works, but as established in Section 10, b-value similarity alone is insufficient for deciding whether zones can be merged. Zones with identical b-values but different area-normalised seismicity densities still require separate treatment in the source model: merging them would misrepresent the spatial distribution of seismicity even though the magnitude-frequency slope is the same. The full aggregation tests found that 68% of adjacent zone pairs are distinguishable when rates are included, compared to only 16% for b-value alone.

To address this, the hierarchical merging is repeated using a joint similarity measure that requires both b-value overlap *and* rate density overlap. The joint similarity score is the minimum of the b-value similarity and the rate density similarity: both must be high for a merge to proceed. Rate density is defined as λ_n / A (events per year per km²), with uncertainty propagated from the posterior σ_{λ_n}.

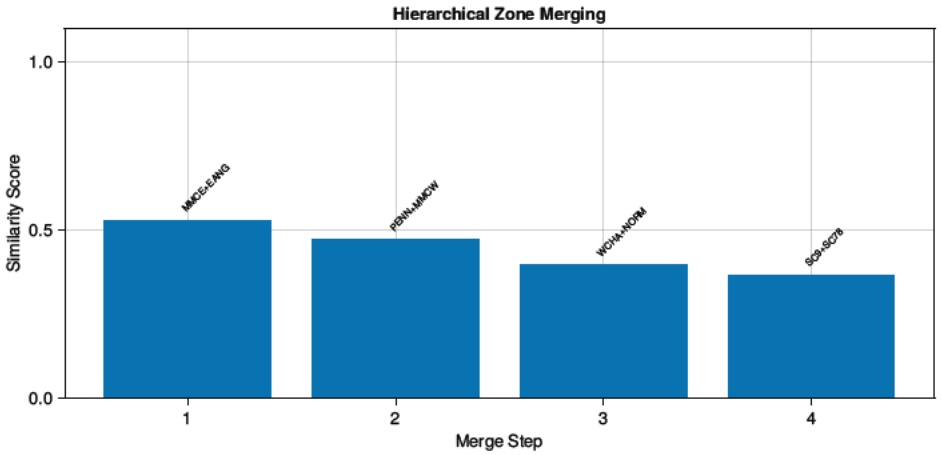

Figure 13.6: Joint similarity scores (minimum of b-value and rate density similarity) for hierarchical zone merging (Mw ≥ 3.0). Compared to Figure 13.3, substantially fewer zone pairs exceed the similarity threshold when rate density differences are also required to be small.

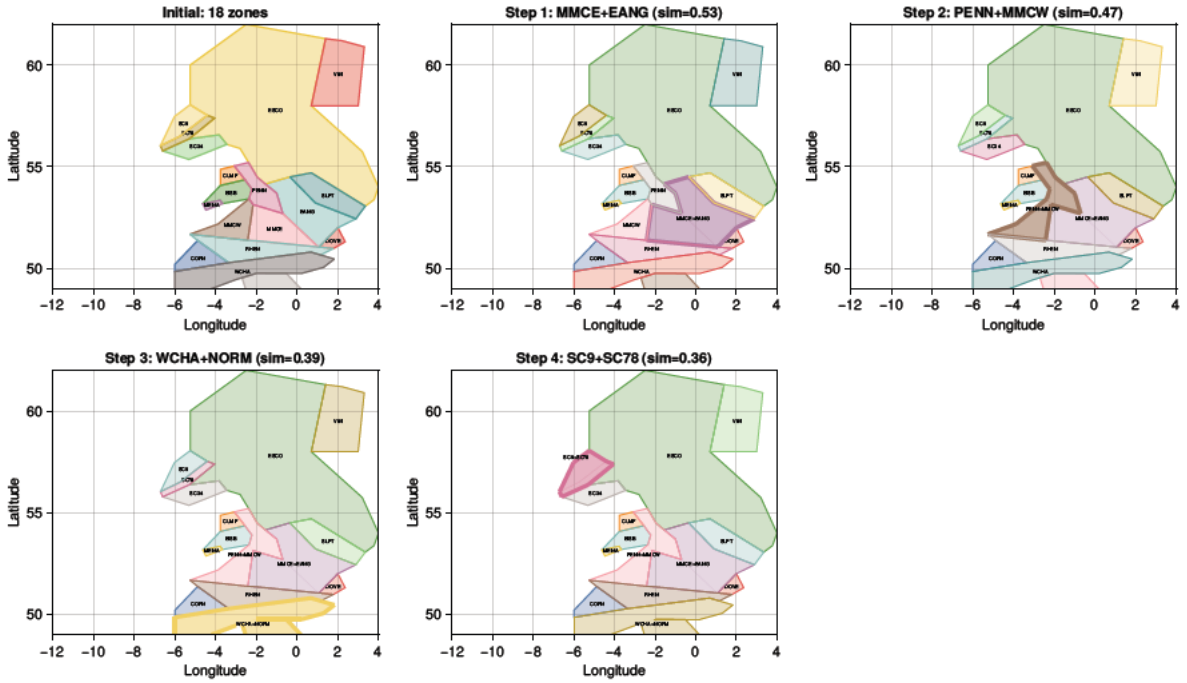

Figure 13.7: Evolution of zone configuration through the merging sequence using joint b-value and rate density similarity. Fewer merges survive compared to the b-value-only analysis (Figure 13.4), preserving more of the existing zone structure.

The joint similarity measure produces a different merge sequence from the b-value-only analysis, because the greedy algorithm always selects the most similar remaining pair. The four surviving merges are:

Table 13.4: Zone merging sequence using joint b-value and rate density similarity (Mw ≥ 3.0). Rate densities ρ are in events per year per 10^4 km^2. Only pairs with similar b-values *and* similar rate densities are merged.

Merge	Zone 1	Zone 2	Similarity	ρ_1	ρ_2
1	MMCE	EANG	0.53	0.022	0.014
2	PENN	MMCW	0.47	0.102	0.083
3	WCHA	NORM	0.39	0.011	0.020
4	SC9	SC78	0.36	0.109	0.170

Notably, some merges from the b-value-only analysis (Table 13.2) no longer occur. MENA+EISB is blocked because MENA's high seismicity density (reflecting its small area and high activity) differs substantially from EISB's. SLPT+EANG is blocked for similar reasons. Conversely, new pairings emerge: PENN+MMCW and SC9+SC78 have both similar b-values and similar rate densities.

The comparison between the two analyses is:

Table 13.5: Merging results at Mw ≥ 3.0 comparing b-value-only and joint (b-value + rate density) similarity. Including rate density changes both the number and identity of merges.

Similarity Measure	Merges	Final Zones	Threshold	Comment
b-value only	5	13	Mw ≥ 3.0	Proof of concept
b-value + rate density	4	14	Mw ≥ 3.0	More realistic

13.3.5 Prior Sensitivity

The choice of prior on the b-value slope parameter β significantly affects both point estimates and uncertainty bounds. Table 13.6 compares results using a weak prior (σ_β = 0.5, data-driven) versus UK practice (σ_β = 0.2, equivalent to the PMLM weight W = 25).

Table 13.6: Prior sensitivity for selected zones. The UK practice prior reduces uncertainty by 40% but also pulls b-values toward b=1.0, homogenising spatial variation.

Zone	N	b (weak)	b (UK)	σ_b reduction
CORN	8	1.04 ± 0.15	1.01 ± 0.08	47%
WCHA	8	0.89 ± 0.15	0.97 ± 0.08	47%
MMCW	23	0.70 ± 0.11	0.87 ± 0.07	36%
PENN	22	0.86 ± 0.11	0.95 ± 0.07	36%
SLPT	17	1.14 ± 0.14	1.05 ± 0.08	43%
VIKI	27	1.14 ± 0.12	1.06 ± 0.07	42%

13.3.6 Discussion

The prior sensitivity analysis reveals an important trade-off. The stronger UK practice prior (PMLM-equivalent) approximately halves posterior uncertainty but simultaneously

homogenises b-value estimates by pulling them toward the prior mean of b = 1.0. For example:

- MMCW: b changes from 0.70 ± 0.11 to 0.87 ± 0.07
- SLPT: b changes from 1.14 ± 0.14 to 1.05 ± 0.08

This has implications for zone merging decisions. With the stronger prior, zones appear more similar (all closer to b = 1.0), potentially increasing mergers. With weaker priors, genuine spatial variation is preserved but uncertainty is larger.

The comparison between b-value-only merging (Section 13.3.3) and joint b-value + rate density merging (Section 13.3.4) is instructive. Using b-value similarity alone, 5 of 18 zone pairs were merged (reducing to 13 zones). When rate density differences are properly accounted for, fewer merges survive. This demonstrates that the existing zonation is better-supported than b-value analysis alone suggests: zones differ primarily in *how much* seismicity they produce, not in *how* that seismicity is distributed across magnitudes. The full aggregation tests from Section 10 provide independent confirmation of this conclusion, finding that 68% of adjacent pairs are distinguishable when rates are included versus only 16% for b-value alone.

This is an important practical finding. It means that the main role of zone boundaries in the current UK source model is to capture spatial variation in seismicity *density*, not in the shape of the magnitude-frequency distribution. Hierarchical models that share a common (or partially pooled) b-value while allowing zone-specific rates may therefore provide a more parsimonious representation of UK seismicity.

This approach is conservative in that it only merges within the existing framework rather than questioning the framework itself.

13.4 Synthesis: What the Approaches Reveal

The analyses provide complementary insights:

- **K-means clustering** shows that seismicity data alone would suggest different boundaries from those currently adopted—only 71-76% of events have consistent zone assignment between data-driven and expert-defined approaches
- **Hierarchical merging (b-value only)** shows that when zones are compared on b-value alone, 5 out of 18 calibratable zone pairs (28%) can be merged without statistical loss
- **Hierarchical merging (full similarity)** shows that when rate density differences are also required to be small, substantially fewer merges survive—confirming that zone boundaries are primarily justified by differences in seismicity density rather than magnitude-frequency slope

The specific findings reinforce each other:

1. **Data coverage is highly variable**: 4 of 22 UK zones (BALA, SC1M, IREL, PASC) have insufficient data at Mw ≥ 3.0 for independent b-value calibration. These zones have no data-driven basis for their current boundaries—parameters must be borrowed from adjacent zones or regional estimates.
2. **b-values are largely homogeneous**: The zones that can be calibrated show substantial overlap in b-value posteriors. The b-value-only merging reduced 18 zones to 13, and the full aggregation tests found only 16% of zone pairs distinguishable by b-value.
3. **Rate density is the primary discriminant**: When area-normalised seismicity rates are included, most zone boundaries are justified. The full aggregation tests found 68% of pairs distinguishable, and the joint-similarity merging blocks most of the b-value-only merges. This points toward hierarchical models with shared b-values and zone-specific rates as a natural model structure.
4. **Prior choice strongly affects conclusions**: The UK practice prior (equivalent to PMLM with W = 25) reduces posterior uncertainty by ~40% but homogenises b-value estimates. This masks genuine spatial variation and can affect which zones appear "distinguishable."

Together, these analyses suggest that current UK zonation represents a specific set of expert choices that could reasonably have been made differently. This is not a criticism of past practice – geological and geophysical reasoning provides valuable context that purely statistical approaches ignore. However, it highlights genuine epistemic uncertainty that is currently unacknowledged in hazard calculations.

Recommendations for Future Practice:

1. **Acknowledge zonation uncertainty**: Zone boundary choices represent significant epistemic uncertainty that is currently unquantified. Approximately 25-30% of events would be classified differently under data-driven vs expert-defined approaches.
2. **Include alternative zonations in logic trees**: Future hazard assessments should consider including both detailed (current 22-zone) and merged configurations as epistemic alternatives, rather than treating boundaries as fixed.
3. **Document prior choices and their effects**: The UK practice prior (PMLM-equivalent) reduces apparent uncertainty by 40% while homogenising b-values. This trade-off should be explicitly acknowledged and sensitivity analyses performed.
4. **Reconsider zone distinctions**: 4 zones have no data support for independent calibration. Using b-value similarity alone, 5 additional zone pairs appear statistically indistinguishable — but when rate density is also considered, most boundaries are justified. This confirms that zones differ primarily in seismicity density, not magnitude-frequency slope, and supports hierarchical models with shared b-values and zone-specific rates.
5. **Consider hierarchical/pooling approaches**: Rather than treating zones as independent, hierarchical models that share information across zones while allowing for spatial variation may provide better uncertainty characterisation.

13.5 Future Extensions

A more complete treatment of flexible source boundaries would build upon the proof-of-concept analyses presented here in several directions:

- **Alternative clustering algorithms**: As noted in the limitations discussion, DBSCAN can discover arbitrarily-shaped clusters and identify outlier events that do not belong to any cluster, while Gaussian mixture models provide probabilistic zone membership rather than hard boundaries. These would better capture the elongated seismicity patterns associated with fault zones and rift margins.
- **Magnitude-weighted clustering**: Weighting events by magnitude or seismic moment would emphasise the events that contribute most to hazard, producing clusters more representative of the underlying tectonic moment release. In the UK context, larger events tend to be historical and may carry *greater* location uncertainty than smaller instrumental events, so magnitude weighting cannot simply be equated with better-located events. A natural approach to handling location uncertainty is to discretise the region into spatial cells and distribute each event's contribution (e.g. moment release rate, computed as the event's

moment divided by the relevant completeness period) across cells in proportion to the event's location probability density, rather than assigning each event to a single point.

- **Three-dimensional zonation**: In principle, incorporating hypocentral depth could add a further dimension to the zonation. However, UK seismicity is entirely crustal with no clearly distinct depth regimes, and the catalogue is already sparse enough that subdividing by depth is unlikely to be productive with the current data.
- **Formal model comparison**: Information criteria (such as the Bayesian Information Criterion or the Widely Applicable Information Criterion) could be used to determine the optimal number of zones — addressing the fundamental question of how many distinct seismic sources the data actually support, rather than fixing K to match current practice.
- **Full flexible-boundary inversion**: The most ambitious extension would treat zone membership as a latent variable estimated jointly with seismicity parameters within the hierarchical model. Each event would have a probability of belonging to each zone, and the zone boundaries would be inferred from the data alongside the b-values and activity rates. This is computationally demanding but would provide a fully Bayesian treatment of zonation uncertainty.
- **Geological prior information as soft constraints**: Rather than ignoring geological context entirely (as K-means does) or treating it as fixed (as current practice does), geological boundaries could be encoded as informative priors on cluster locations — preferring but not requiring that seismicity clusters align with known tectonic features.

These extensions represent substantial methodological development and are deferred to future work.

14 Alternative Occurrence Models: Poisson vs Negative Binomial

14.1 Motivation: Testing the Poisson Assumption

The standard approach to earthquake occurrence modelling assumes that event counts follow a Poisson distribution, where the variance equals the mean. However, as noted in Section 3.3, there is evidence that this equality does not always hold for earthquake occurrence, with the variance frequently exceeding the mean. This is a phenomenon known as overdispersion.

Overdispersion in earthquake catalogues can arise from several sources:

- Residual clustering not fully removed by declustering
- Spatial heterogeneity within analysis regions
- Temporal non-stationarity in seismicity rates
- Magnitude uncertainty effects on binned counts

The negative binomial distribution provides a natural extension to the Poisson that accommodates overdispersion. It introduces a dispersion parameter φ such that:

$$\mathrm{Var}(N) = \mu + \frac{\mu^2}{\varphi} \tag{14.1}$$

As $\varphi \to \infty$, the negative binomial converges to the Poisson. When φ is small, the variance substantially exceeds the mean, indicating significant overdispersion.

This section tests whether the negative binomial provides a better fit to UK seismicity data than the Poisson assumption embedded in standard practice. Two complementary approaches are used: a PMLM-style analysis that mirrors current practice, and a hybrid Binned L5 model that retains the full Full Bayesian treatment of magnitude uncertainties while enabling overdispersion detection.

14.2 PMLM-Style Analysis with Overdispersion

To make the comparison directly relevant to practitioners, the first analysis mirrors the familiar PMLM methodology (Johnston et al., 1994):

- Uses binned magnitude counts with variable completeness periods (exactly as in standard practice)
- Applies the same prior on β (equivalent to PMLM penalty)
- For the Poisson model: standard binned likelihood as in Equation 3.33

- For the Negative Binomial: replaces the Poisson likelihood with $n_i \sim \text{NegBin}(\mu_i, \varphi)$

This formulation allows direct comparison. It essentially says, we'll adopt the same approach as what is done in current practice, but we relax a single assumption related to the mean-variance equality of the Poisson assumption. And, this is the impact when that assumption is relaxed.

The negative binomial likelihood for binned counts is:

$$\mathcal{L}_{\text{NegBin}} = \prod_{i=1}^{I} \text{NegBin}(n_i \mid \mu_i, \varphi) \tag{14.2}$$

where the negative binomial probability mass function (using the mean-dispersion parameterisation) is:

$$P(n \mid \mu, \varphi) = \frac{\Gamma(n+\varphi)}{\Gamma(\varphi) \cdot n!} \cdot \left(\frac{\varphi}{\varphi+\mu}\right)^{\varphi} \cdot \left(\frac{\mu}{\varphi+\mu}\right)^{n} \tag{14.3}$$

14.3 Results: Evidence for Overdispersion

The comparison was performed at three magnitude thresholds using the complete UK catalogue. Table 14.1 shows that the point estimates are nearly identical between models, while Table 14.2 shows the impact on uncertainty quantification.

Table 14.1: Parameter estimates under Poisson and Negative Binomial models. For Mw ≥ 3.0 and Mw ≥ 2.5, point estimates are essentially identical (< 2% difference in b-value, < 6% in rate). At Mw ≥ 2.0, stronger overdispersion leads to larger differences, but the models still agree within uncertainty.

Threshold	**N**	b (Poisson)	b (NegBin)	Difference
Mw ≥ 3.0	217	0.968 ± 0.049	0.969 ± 0.061	< 0.1%
Mw ≥ 2.5	427	1.021 ± 0.031	1.000 ± 0.046	2.1%
Mw ≥ 2.0	1586	1.182 ± 0.019	1.070 ± 0.042	10.5%
Threshold	**N**	λ (Poisson)	λ (NegBin)	Difference
Mw ≥ 3.0	217	2.12 ± 0.17	2.24 ± 0.30	5.7%
Mw ≥ 2.5	427	7.54 ± 0.41	7.59 ± 0.92	0.7%
Mw ≥ 2.0	1586	38.4 ± 1.0	32.6 ± 4.3	15.1%

Table 14.2: Overdispersion diagnostics. The dispersion parameter φ (smaller = more overdispersion), overdispersion factor ($\frac{\text{Var}}{\text{Mean}}$), and ratios of parameter uncertainties between models. All ratios show NegBin uncertainty divided by Poisson uncertainty. The key finding is that rate uncertainty is underestimated by 1.8–4.4× under the Poisson assumption.

Threshold	**N**	φ	**Overdispersion**	σ_λ ratio	σ_b ratio
Mw ≥ 3.0	217	10.6 ± 6.7	2.9×	1.8×	1.2×
Mw ≥ 2.5	427	10.1 ± 5.1	4.5×	2.3×	1.5×
Mw ≥ 2.0	1586	5.3 ± 1.8	30×	4.4×	2.1×

Table 14.1 and Table 14.2 together show that while the point estimates for b-value and activity rate are largely consistent between the Poisson and Negative Binomial models, the uncertainties differ substantially. The negative binomial model reveals that the Poisson assumption leads to a significant underestimation of uncertainty, particularly for the activity rate.

For Mw ≥ 3.0 and Mw ≥ 2.5, the point estimates for b-value differ by less than 2.1%, and activity rates by less than 6%. The models agree on what the parameters are – but they disagree on how confident we should be about those evaluations. At Mw ≥ 2.0, stronger overdispersion leads to ~10% differences in point estimates, though the values still agree within their respective uncertainties.

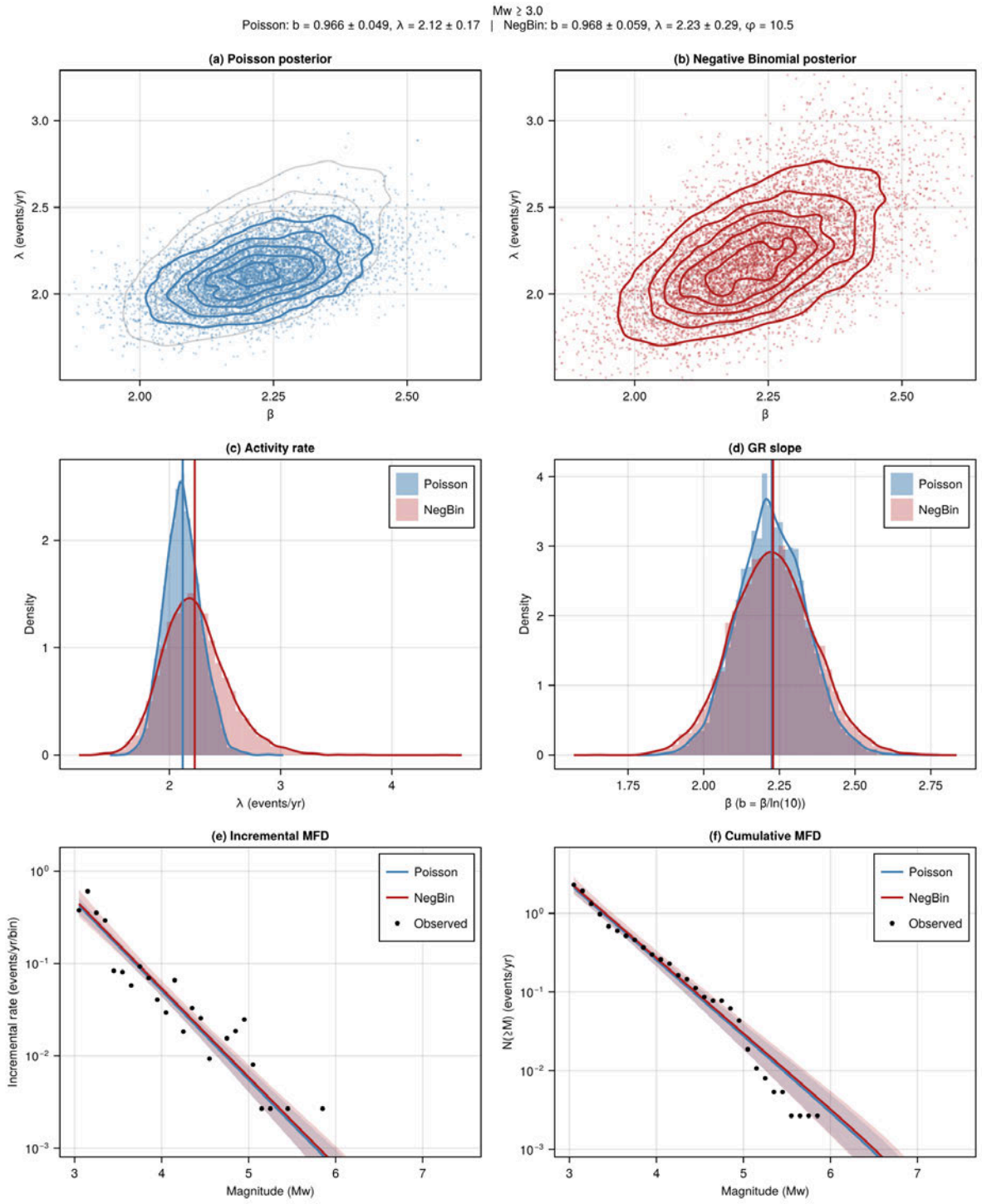

Figure 14.1: Comparison of Poisson and Negative Binomial models for UK seismicity at Mw ≥ 3.0 (N = 217 events). (a) Joint posterior for λ and β: both models give similar central values, but NegBin samples show greater dispersion. (b) Marginal posterior for activity rate λ: note the wider NegBin distribution (1.8× larger standard deviation). (c) Marginal posterior for GR slope β (equivalently, $b = \frac{\beta}{\ln(10)}$). (d) Magnitude-frequency distribution with 95% credible intervals: observed data (black circles) are cumulative rates N(≥M) computed using appropriate completeness-adjusted observation times. Both models fit the mean equally well, but NegBin gives wider uncertainty bands.

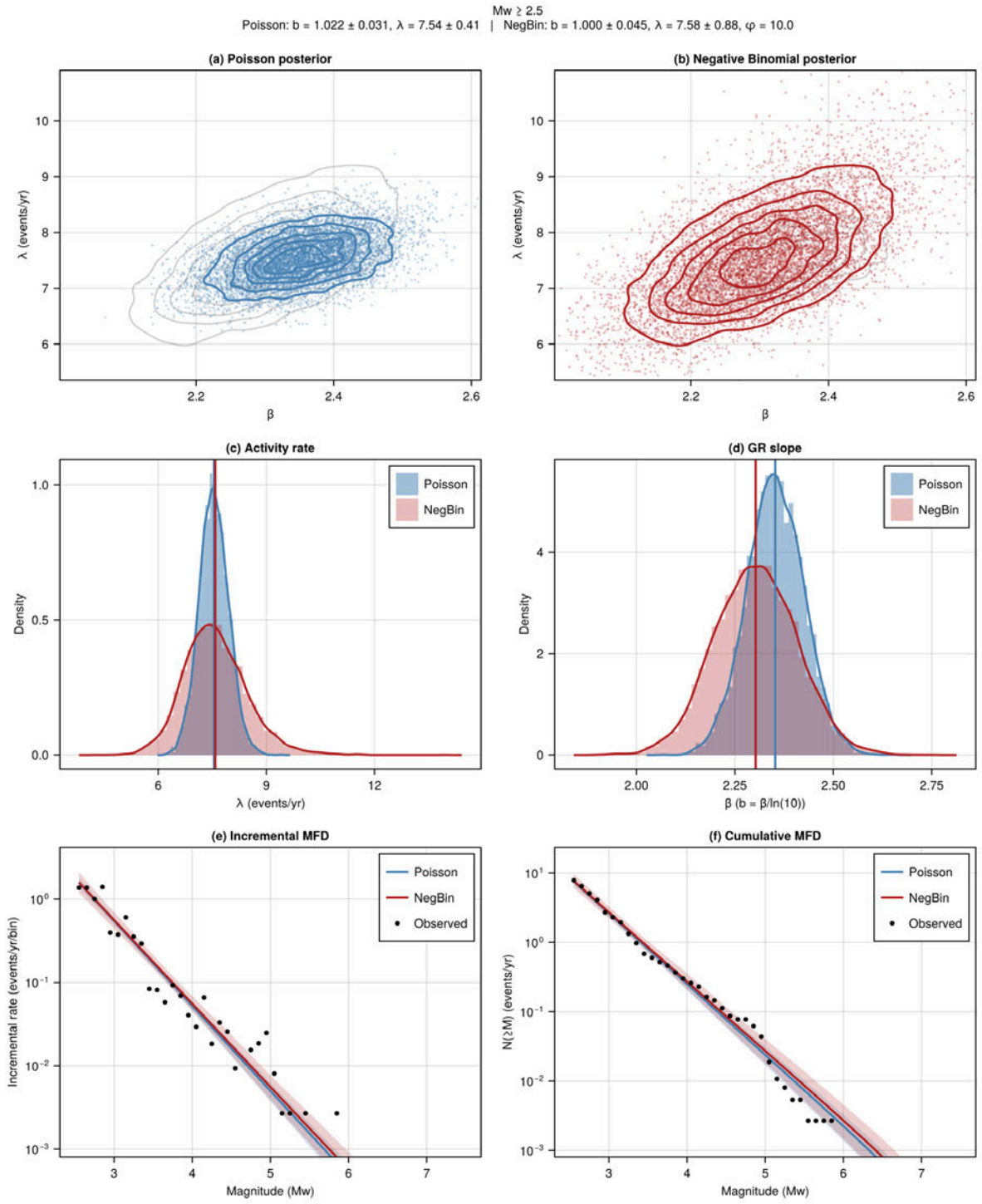

Figure 14.2: Comparison of Poisson and Negative Binomial models at Mw ≥ 2.5 (N = 427 events). The pattern is similar to Figure 14.1 but overdispersion is more pronounced. The NegBin rate uncertainty (σ_λ) is 2.3× larger than Poisson, indicating that rate uncertainty is substantially underestimated by the Poisson assumption at this threshold.

Figure 14.1 and Figure 14.2 visually illustrate the impact of overdispersion on parameter uncertainty. The negative binomial model produces wider credible intervals for both the activity rate and the GR slope, reflecting the additional variability in the data that the Poisson model fails to capture. The upper rows of these figures show the joint posterior distributions,

with the other posterior shown in grey for reference. For example, in the upper left panels the blue contours show the joint posterior for the Poisson model, while the grey contours show the NegBin posterior. The NegBin posterior is more dispersed, indicating greater uncertainty about the parameters. The marginal posteriors for λ and β further highlight the increased uncertainty under the negative binomial model.

14.4 Confirmation under the Full Bayesian Model

A potential objection to the PMLM-based overdispersion analysis is that PMLM ignores magnitude measurement uncertainty: observed magnitudes are treated as exact when computing expected bin counts. If measurement uncertainty causes events to scatter between bins, this could inflate the apparent variance and produce a spurious overdispersion signal. To test this, a hybrid "Binned L5" model was developed that retains the full Full Bayesian (L5) treatment of magnitude uncertainties while enabling overdispersion detection through binned counts.

The key insight is that the unbinned L5 model has only one count observation (the total N), which provides no information about the dispersion parameter φ. The Binned L5 model resolves this by binning events by their observed M_L values (giving ≈ 25–46 bins), while computing expected counts per bin using the full L5 machinery. Specifically, the expected count in M_L bin $[m_l^{\mathrm{lo}}, m_l^{\mathrm{hi}})$ is:

$$\mu_j = \lambda_0 \int_{m_w^{\mathrm{floor}}}^{m_w^{\max}} T(m_w) \cdot f_{\mathrm{GR}}(m_w \mid \beta) \cdot P_{\mathrm{bin}}(j \mid m_w)\, \mathrm{d}m_w \tag{14.4}$$

where P_{bin} replaces the selection probability of the standard L5 model:

$$P_{\mathrm{bin}}(j \mid m_w) = \Phi\left(\frac{m_l^{\mathrm{hi}} - m_l^{\mathrm{true}}(m_w)}{\sigma_{\mathrm{total}}}\right) - \Phi\left(\frac{m_l^{\mathrm{lo}} - m_l^{\mathrm{true}}(m_w)}{\sigma_{\mathrm{total}}}\right) \tag{14.5}$$

Here $m_l^{\mathrm{true}}(m_w)$ is the Grünthal conversion and σ_{total} combines measurement uncertainty, conversion uncertainty, and rounding in quadrature — exactly as in the standard L5 model. The integral in Equation 14.4 accounts for variable completeness through $T(m_w)$ and uses the same quadrature scheme as the unbinned L5.

This formulation has the same physics as the unbinned L5, but its binned count structure provides ≈ 25–46 independent observations for estimating φ. As a validation check, the Binned L5 Poisson model was compared against the unbinned L5 Poisson: the two agree to within 0.5% in b-value and 0.3% in activity rate at all thresholds, confirming that the binning does not distort the L5 parameter estimates.

Table 14.3: Comparison of dispersion parameter φ between PMLM and Binned L5 models. The Binned L5 consistently estimates *smaller* φ (more overdispersion) than PMLM. This demonstrates that magnitude uncertainty does not explain the overdispersion — it obscures additional overdispersion that the L5 treatment reveals.

Threshold	N	**ML bins**	φ (PMLM)	φ (Binned L5)	φ ratio
Mw ≥ 3.0	217	34	10.5 ± 6.6	5.0 ± 2.5	0.48
Mw ≥ 2.5	427	40	10.0 ± 5.2	4.2 ± 1.6	0.42
Mw ≥ 2.0	1586	46	5.3 ± 1.8	2.8 ± 0.8	0.53

Table 14.4: Parameter estimates under Binned L5 Poisson and Negative Binomial models. The Binned L5 Poisson values are consistent with the unbinned L5 Poisson (validation). Under the NegBin, rate uncertainty increases by 2–5× while b-value uncertainty roughly doubles.

Threshold	b (Binned L5 Pois.)	b (Binned L5 NegBin)	λ (Binned L5 Pois.)	λ (Binned L5 NegBin)
Mw ≥ 3.0	0.855 ± 0.043	0.931 ± 0.074	1.60 ± 0.11	1.73 ± 0.25
Mw ≥ 2.5	0.892 ± 0.027	0.950 ± 0.053	5.03 ± 0.24	5.44 ± 0.71
Mw ≥ 2.0	1.116 ± 0.020	1.055 ± 0.046	25.4 ± 0.7	24.2 ± 3.4

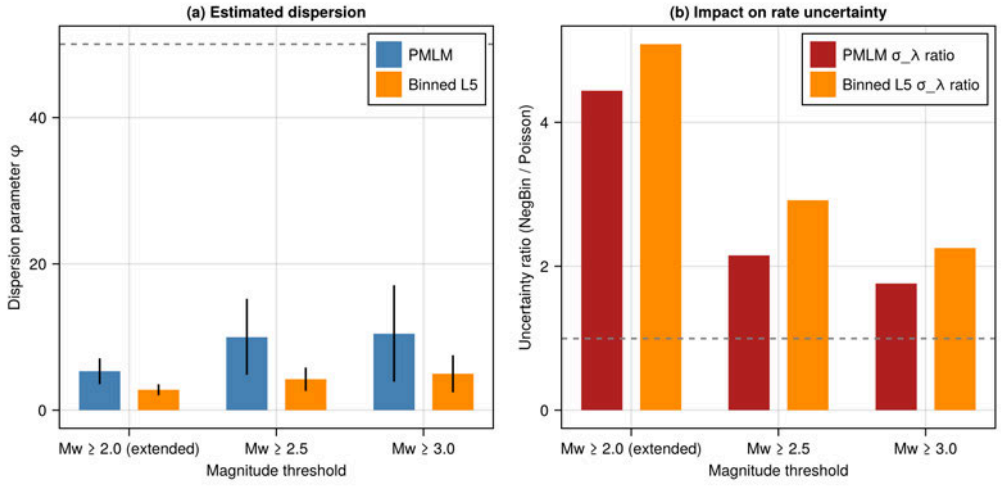

Figure 14.3: Summary of overdispersion findings across PMLM and Binned L5 models. (a) Estimated dispersion parameter φ at each threshold: the Binned L5 (orange) consistently shows smaller φ (more overdispersion) than PMLM (blue), demonstrating that the finding is robust to the treatment of magnitude uncertainties. (b) Impact on rate uncertainty: the NegBin gives 2–5× larger rate uncertainty across both model families.

Table 14.3 presents the central result: at every threshold, the Binned L5 model estimates a *smaller* φ than PMLM, indicating *more* overdispersion when magnitude uncertainties are properly accounted for. This is the opposite of what would be expected if PMLM's overdispersion were an artefact of ignoring measurement uncertainty. The φ ratio (Binned L5 / PMLM) is consistently around 0.4–0.5, meaning the L5 treatment roughly doubles the apparent overdispersion.

The interpretation is that PMLM's binning in Mw space, without accounting for the scatter of events between bins due to measurement uncertainty, slightly *smooths* the bin counts — making them appear more regular than they truly are. The L5 model, by correctly modelling the observation process, reveals the full extent of the count variability.

14.5 Interpretation

14.5.1 Overdispersion Increases at Lower Thresholds

The dispersion parameter φ decreases (indicating more overdispersion) as the magnitude threshold is lowered. Under PMLM: $\varphi \approx 10.5$ (Mw ≥ 3.0), 10.0 (Mw ≥ 2.5), and 5.3 (Mw ≥ 2.0). Under the Binned L5: $\varphi \approx 5.0$, 4.2, and 2.8 respectively. Both models show the same pattern, with the Binned L5 consistently reporting stronger overdispersion.

The PMLM values at Mw ≥ 3.0 and Mw ≥ 2.5 are similar, indicating comparable levels of overdispersion at these thresholds. A more marked increase occurs at Mw ≥ 2.0, where φ roughly halves under both models.

This pattern is expected because:

1. Lower thresholds include more small events, which exhibit stronger clustering
2. Spatial heterogeneity becomes more apparent with larger sample sizes
3. Catalogue completeness becomes more variable at lower magnitudes

However, the sharp increase at Mw ≥ 2.0 may partly be a catalogue artefact. The completeness thresholds adopted for Mw ≥ 3.0 are conservative and well-established, and Mosca et al. (2022) suggests that Mw ≥ 2.5 is also reasonable. Extending to Mw ≥ 2.0 is more speculative: the network detection capability at this level varies spatially, and events near recording stations are more likely to be catalogued than those in poorly instrumented areas. This could introduce spatial clustering related to station density rather than tectonic structure, which would manifest as additional overdispersion. The extent to which the lower φ at Mw ≥ 2.0 reflects genuine seismicity clustering versus incomplete catalogue coverage is therefore difficult to separate.

14.5.2 Rate Uncertainty is Systematically Underestimated

The most significant finding is that activity rate uncertainty is underestimated by factors of 2–4× when using the Poisson assumption. The GR slope (β or b-value) is less affected, with NegBin uncertainties 30–100% larger than Poisson.

This has direct implications for hazard assessment:

- Annual probabilities of exceedance inherit rate uncertainty
- If rate uncertainty is underestimated, hazard confidence intervals are too narrow
- The P84 (84th percentile) hazard is likely to be systematically too low

14.5.3 Model Selection

The log-likelihood strongly favours the negative binomial at all thresholds:

Table 14.5: Log-likelihood comparison. Positive ΔLL indicates NegBin is preferred. The difference is especially large at Mw ≥ 2.0 where overdispersion is most pronounced.

Threshold	LL (Poisson)	LL (NegBin)	ΔLL
Mw ≥ 3.0	−78.9	−73.4	+5.5
Mw ≥ 2.5	−103.5	−93.8	+9.7
Mw ≥ 2.0	−277.4	−126.4	+151.0

Even accounting for the additional parameter in the negative binomial model, the improvement is substantial—particularly at Mw ≥ 2.0 where the log-likelihood difference of 151 units represents overwhelming evidence against the Poisson assumption. Figure 14.3 summarises the dispersion estimates and uncertainty impact across both PMLM and Binned L5 model families.

14.6 Implications for Practice

The Poisson assumption in standard seismicity analysis systematically underestimates activity rate uncertainty. This finding is robust to the treatment of magnitude uncertainties.

- For Mw ≥ 3.0 (typical operational threshold): rate uncertainty is underestimated by 80%
- For Mw ≥ 2.5: rate uncertainty is underestimated by 120%
- Point estimates (b-value, mean rate) are unaffected: only uncertainty quantification is impacted
- The Binned L5 model, which accounts for magnitude uncertainties, conversion, and selection effects, confirms and strengthens the overdispersion finding (φ roughly halved compared to PMLM)

The overdispersion likely reflects genuine variability in UK seismicity rates – from residual clustering, spatial heterogeneity, or temporal non-stationarity – that cannot be captured by a homogeneous Poisson process.

For hazard applications:

1. **Consider using negative binomial** for uncertainty quantification, particularly when lower magnitude thresholds are used
2. **Inflate Poisson-derived uncertainties** as a first-order correction — multiplying by 1.5–2× would provide approximate consistency with NegBin estimates

3. **Interpret P84 estimates with caution** when they are based on Poisson assumptions, particularly for sites where small-to-moderate earthquakes dominate the hazard

The negative binomial model can be viewed as a gamma-Poisson mixture, where unobserved heterogeneity in the underlying rate manifests as overdispersion. This interpretation suggests that overdispersion reflects genuine uncertainty about the "true" rate that cannot be resolved with the available data.

15 Conclusions and Recommendations

This report addresses three themes relating to the calibration of magnitude-frequency distributions for seismic hazard assessment at UK nuclear sites. Theme A examines optimal calibration methods, focusing on how magnitude uncertainties, rounding effects, and selection bias affect the estimation of Gutenberg-Richter parameters – and whether the penalized maximum likelihood method (PMLM) currently used in UK practice adequately accounts for these effects. Theme B investigates how epistemic uncertainty in the GR parameters should be characterised and propagated through logic trees for hazard calculations. Theme C explores whether information can be shared across source zones through hierarchical models and external constraints. An additional investigation tests the Poisson assumption underlying all standard approaches.

The overarching conclusion is that current UK practice, while producing reasonable point estimates of GR parameters, systematically underestimates epistemic uncertainty through multiple mechanisms: ignoring magnitude uncertainty (Theme A), using discretization schemes that suppress distributional tails (Theme B), treating zones as independent when they share prior information and physical constraints (Theme C), and assuming Poisson occurrence when the data exhibit significant overdispersion. These effects compound in hazard calculations, leading to regulatory fractiles (e.g., P84) that are likely too low.

15.1 Key Findings

15.1.1 Theme A: Optimal Calibration

1. The penalized MLE method (PMLM; (Johnston et al., 1994)) is mathematically equivalent to maximum a posteriori (MAP) estimation with a Normal prior on β. The UK standard weight $W = 25$ corresponds to a prior standard deviation of $\sigma_b \approx 0.087$.
2. PMLM and all other standard methods (Weichert MLE, Aki MLE) ignore magnitude measurement uncertainty. This produces systematic bias: approximately 5% in the b-value and 12% in the activity rate under typical UK conditions ($\sigma_{M_L} \approx 0.3$, buffer ≈ 1.0 magnitude units).
3. The Full Bayesian Bayesian model (L5), which marginalises over latent true magnitudes and explicitly models selection effects, reduces bias to $< 2\%$ and produces well-calibrated posterior uncertainties. Monte Carlo validation across 100 synthetic catalogues confirms that nominal 95% credible intervals contain the true parameter values 91% of the time for b and 97% for λ.

4. The N^*/M^* correction method (Tinti et al., 2005) provides a computationally cheaper alternative that addresses the leading-order magnitude uncertainty effects. It does not model selection or variable per-event uncertainties, but for well-buffered catalogues its bias is typically within 5% of the Full Bayesian.
5. Current UK practice achieves approximately a 3σ buffer between completeness and analysis thresholds, which provides substantial (but not complete) protection against Eddington bias. Reducing this buffer – a natural temptation for data-sparse zones – erodes this protection and requires explicit treatment of magnitude uncertainty.

15.1.2 Theme B: Epistemic Uncertainty and Logic Trees

1. The correlation between rate and slope depends on the reference magnitude: $\rho(\lambda_n, \beta) \approx +0.4$ (positive) at the lower magnitude bound, but $\rho(\lambda(m_{\text{ref}}), \beta) \approx -0.7$ to -0.85 (strongly negative) at hazard-relevant magnitudes. It is the negative correlation that matters for hazard calculations; ignoring it overestimates uncertainty.
2. Standard Miller-Rice discretization, used to construct 3-point logic tree branches, produces 7–8% errors in the P84 hazard fractile. This is not a minor numerical issue – it reflects a systematic bias in how distributional tails are represented. An optimised Heavy-Tail scheme achieves errors of $\approx 1.5\%$.
3. The PMLM prior ($\sigma_b \approx 0.087$) systematically suppresses zone-level uncertainty by 40–80% compared to the variability observed across UK source zones. This means that the logic tree branches derived from PMLM understate the true range of epistemic uncertainty.
4. 68% of UK source zones are statistically indistinguishable from the national average b -value. Apparent between-zone variability is largely an artefact of small sample sizes rather than genuine tectonic differences.
5. Existing approaches to constructing logic trees for the GR parameters lack a rigorous mathematical basis. Current practice selects discrete parameter values and assigns weights using ad hoc methods (often expert judgment or simple quantile mapping) that do not properly preserve the joint distribution, its correlation structure, or the moments required for unbiased hazard estimation. The discretization schemes investigated here provide a principled alternative with quantifiable approximation errors.

15.1.3 Theme C: Joint Inversion and Multi-Zone Analysis

1. Hierarchical Bayesian models enable partial pooling of information across source zones. Shrinkage stabilises estimates for data-poor zones while allowing genuinely anomalous zones to retain distinct parameter values.
2. External constraints (particularly seismic moment release rates) can further regularize inversions and prune physically implausible logic tree branches. UK NSHM 2020 moment release is broadly consistent with geodetic constraints.

3. Data-driven zonation (K-means clustering) assigns 24–29% of events to different zones than expert-defined boundaries, indicating substantial unquantified epistemic uncertainty in current zone geometries.
4. Many current zone distinctions are not statistically supported. Hierarchical merging reduces 18 zones to 13 without statistical loss, suggesting that simplified zonations may better represent the information actually available.

15.1.4 Alternative Occurrence Models

1. UK seismicity exhibits significant overdispersion ($\mathrm{Var} > \mathrm{Mean}$ for event counts) at all magnitude thresholds tested. Under PMLM, $\varphi \approx 5$–11; under the Binned L5 model (which retains full Full Bayesian treatment of magnitude uncertainties), $\varphi \approx 3$–5. The finding is robust to methodology and, if anything, is strengthened when magnitude uncertainties are properly accounted for.
2. Point estimates (b-value, mean rate) are essentially identical between Poisson and negative binomial models – only the uncertainty changes. Rate uncertainty under the negative binomial is 1.8–4.5× larger than under the Poisson assumption.
3. Log-likelihood strongly favours the negative binomial, especially at Mw $\geq$ 2.0 (ΔLL = +151 for PMLM, +276 for Binned L5), representing overwhelming statistical evidence against the Poisson assumption.

15.2 Recommendations for Practice

15.2.1 Magnitude Uncertainty

- **R1:** Magnitude measurement uncertainty must be accounted for in GR parameter estimation. Methods that ignore it (Weichert MLE, Aki MLE, standard PMLM) produce systematic bias in both the b-value and activity rate. The magnitude of this bias depends on the buffer between the completeness threshold and the analysis threshold, but is non-negligible under typical UK conditions.
- **R2:** The Full Bayesian Bayesian model (L5) is the recommended approach for applications requiring rigorous uncertainty quantification, including nuclear safety assessments. It marginalises over latent true magnitudes, accounts for per-event measurement uncertainties, models selection effects explicitly, and handles variable completeness periods. It should be used regardless of sample size.
- **R3:** Where computational cost or workflow constraints preclude the Full Bayesian, the N^*/M^* correction method provides a reasonable alternative for well-buffered catalogues. It does not replace full Bayesian inference but captures the leading-order magnitude uncertainty effects.

- **R4:** The PMLM method currently used in UK practice is systematically biased for any catalogue where the buffer between completeness and analysis thresholds is less than approximately 3σ. Even with adequate buffering, PMLM fails to propagate magnitude uncertainty into the posterior, leading to underestimated parameter uncertainty. It should not be used as the sole basis for regulatory hazard assessments without correction or supplementary analysis.

15.2.2 Logic Tree Construction

- **R5:** Existing approaches to constructing logic trees for GR parameters lack a rigorous mathematical basis and should be replaced by methods with quantifiable approximation properties. Standard practice of selecting parameter values and assigning weights by expert judgment does not preserve the joint distribution, its correlation structure, or the moments required for unbiased hazard estimation.
- **R6:** The correlation between rate and slope must be preserved when discretizing the joint posterior for logic trees. At hazard-relevant reference magnitudes, the correlation is strongly negative ($\rho \approx -0.7$ to -0.85). Ignoring this correlation – as is standard in current practice – leads to incorrect propagation of epistemic uncertainty.
- **R7:** Miller-Rice discretization should not be used for P84 or other regulatory fractile estimation. It systematically underestimates tail probabilities by 7–8%. The Heavy-Tail scheme ($z = \pm 1.03$, $w = [0.47, 0.06, 0.47]$) or optimised 2×3 asymmetric grids should be used instead, with approximation errors below 2%.
- **R8:** The PMLM prior ($W = 25$, $\sigma_b \approx 0.087$) acts as a strong constraint that suppresses between-zone variability by 40–80%. For fractile-sensitive applications, the prior standard deviation should be increased to $\sigma_b = 0.15$–0.20, or hierarchical Bayesian models should be used to let the data inform the appropriate level of between-zone variability.

15.2.3 Multi-Zone Analysis and Source Zonation

- **R9:** When multiple source zones are analysed, hierarchical models should be considered. Partial pooling produces more robust estimates for data-poor zones than either independent estimation (high variance) or forcing a common b-value (high bias).
- **R10:** Moment release constraints should be applied to prune physically implausible logic tree branches. Unconstrained logic trees can imply moment release rates that are inconsistent with geodetic observations by orders of magnitude.
- **R11:** Correlations between zone parameters arising from shared priors, moment constraints, and boundary uncertainty must be accounted for in logic tree construction. Standard practice treats zone parameters as independent, which is inconsistent with all three of these correlation sources.

- **R12:** Zone boundary uncertainty is a source of epistemic uncertainty that is currently unquantified in UK practice. Zone distinctions should be justified quantitatively rather than inherited from precedent. Where adjacent zones are statistically indistinguishable, merging should be considered.
- **R13:** For comprehensive uncertainty quantification, alternative zonations (not just alternative parameters within fixed zones) should be included as logic tree branches.

15.2.4 Occurrence Model

- **R14:** The Poisson assumption for earthquake occurrence, embedded in all standard methods, systematically underestimates activity rate uncertainty by factors of 2–4×. This finding is confirmed under both PMLM and the Full Bayesian L5 framework (via the Binned L5 model), ruling out magnitude uncertainty as an explanation. The negative binomial distribution should be considered for uncertainty quantification, particularly when lower magnitude thresholds ($M_w < 3.0$) are used.
- **R15:** As a first-order correction when using Poisson-based methods, rate uncertainties should be inflated by a factor of 1.5–2× to approximate the additional variance from overdispersion.
- **R16:** P84 hazard estimates based on Poisson assumptions should be interpreted with caution. Overdispersion in the rate propagates directly to hazard confidence intervals and regulatory fractiles.

15.3 Future Work

1. Implement soft completeness periods to replace the current hard-threshold model
2. Joint inference including maximum magnitude as a model parameter
3. Full hierarchical analysis of all UK source zones with partial pooling
4. Integration of moment rate constraints directly within the hierarchical framework
5. Spatial correlation models for zone parameters
6. Periodic validation of magnitude conversion relationships as network instrumentation improves
7. Physics-based magnitude conversion incorporating Deichmann (2017) asymptotic constraints
8. Heteroscedastic conversion uncertainty models
9. Advanced clustering methods for data-driven zonation (DBSCAN, Gaussian mixtures, spatial point processes)
10. Flexible-boundary inversion where zone membership is treated as a latent variable
11. Formal treatment of zone boundary uncertainty in logic tree construction
12. Posterior predictive assessment of overdispersion under the Full Bayesian model

Appendix: Code and Data Availability

All methods described in this report have been implemented as publicly available, open-source software in two companion packages:

- **`SeismicityAnalysis.jl`** (Julia): https://github.com/pstafford/SeismicityAnalysis.jl
- **`seismicity_analysis`** (Python): https://github.com/pstafford/seismicity_analysis

Both packages provide equivalent functionality, enabling practitioners to use whichever language best fits their workflow. The Julia package additionally includes a standalone R/rstan interface for the Full Bayesian L5 model, so that practitioners who prefer R can use the Bayesian methods directly without a Julia installation. As this work was funded by the Office for Nuclear Regulation, the code is freely available for use; users are asked to attribute the software to this project when using it in published work.

The packages cover the three research themes of this project and provide the following capabilities:

Theme A — Optimal Calibration. Classical maximum likelihood methods (Weichert (1980) MLE; Johnston et al. (1994) penalised MLE) are provided with full covariance matrices, together with the bias correction methods of Tinti & Mulargia (1985) and Rhoades (1996), and the N*/M* methods from NUREG-2115. The Full Bayesian L5 model is implemented with Stan backends in both packages (via StanSample.jl in Julia and CmdStanPy in Python), and additionally with a Turing.jl backend in Julia. The Turing implementation uses a fully marginalised likelihood with numerical quadrature (no latent magnitude parameters), while the Stan implementation uses latent variables and supports both constant and variable completeness. Synthetic catalogue generation with the full forward simulation pipeline (true magnitudes → conversion → measurement noise → rounding → selection) enables Monte Carlo validation.

Theme B — Epistemic Uncertainty. Five discretization schemes are provided for constructing logic tree branches from continuous posterior distributions: Miller-Rice, Heavy-Tail, ESM, EPT, and equal-weight. All schemes support bivariate discretization of the joint (λ, b) posterior, preserving the correlation structure via conditional distributions. The mean–fractile trade-off identified in this report (whereby Miller-Rice discretization underestimates regulatory fractiles by 20–40%) is directly demonstrable with the packages.

Theme C — Joint Inversion. Four hierarchical models are implemented for multi-zone analysis: independent (no pooling), shared b-value, partial pooling with hyperpriors on μ_β and σ_β, and a correlated model with population-level (λ, β) correlation. Negative binomial models for overdispersion are provided in both binned PMLM-style and binned L5 variants.

Analytical seismic moment release rate integration over the doubly-bounded Gutenberg–Richter distribution, together with log-normal importance sampling, enables moment release constraints to be applied to logic trees and posterior samples.

The Julia package uses Julia's package extension mechanism so that the core functionality (classical methods, magnitude conversion, completeness models, synthetic data generation, discretization, and moment constraints) has no MCMC dependencies. The Bayesian models are loaded automatically when the user imports Turing.jl or StanSample.jl, and plotting functionality is available via CairoMakie.jl. The Python package follows a similar design, with core dependencies limited to NumPy, SciPy, and pandas, while the Stan-based Bayesian models require the optional CmdStanPy dependency.

Table 15.1 summarises the module structure, which is common to both packages.

Table 15.1: Module structure common to both the Julia and Python packages. Stan model files for the L5, hierarchical, and negative binomial models are bundled in a `stan/` subdirectory within each package.

Module	Functionality
`types`	Core data types (`Catalogue`, `GRConfig`, `GRResult`, etc.)
`conversions`	Grünthal (2009) M_L–M_w conversion with magnitude-dependent uncertainty
`completeness`	Completeness models (UK default from Mosca et al. 2020) and observation periods
`catalogue`	BGS catalogue loading, completeness filtering, magnitude binning
`classical`	Weichert MLE, PMLM, Tinti rate correction, Rhoades b-value correction, N*/M*
`full_bayesian`	L5 marginalised thinned Poisson (Stan backend; additionally Turing in Julia)
`hierarchical`	Multi-zone partial pooling (independent, shared-b, hierarchical, correlated)
`negative_binomial`	Overdispersion models (PMLM-style and binned L5 with NegBin counts)
`moment_constraints`	Analytical moment integration, sensitivity analysis, importance sampling
`discretization`	Logic tree construction (Miller-Rice, Heavy-Tail, ESM, EPT, equal-weight)
`synthetic`	Full forward simulation for validation (constant and variable completeness)

Data. The BGS earthquake catalogue (accessed December 2025) is not bundled with either package; users must obtain the catalogue data independently from the British Geological Survey. Source zone definitions follow Mosca et al. (2022), the completeness model follows

Musson & Sargeant (2007) and Mosca et al. (2022), and the magnitude conversion follows Grünthal et al. (2009).

References

Aki, K. (1965). Maximum likelihood estimate of b in the formula log N = a - bM and its confidence limits. *Bulletin of the Earthquake Research Institute, Tokyo University*, *43*, 237–239.

Baker, J., Bradley, B., & Stafford, P. (2021). *Seismic Hazard and Risk Analysis*. Cambridge University Press. https://doi.org/10.1017/9781108425056.005

Bender, B. (1983). Maximum-likelihood estimation of b-values for magnitude grouped data. *Bulletin of the Seismological Society of America*, *73*(3), 831–851.

Bourne, S. J., & Oates, S. J. (2020). Stress-Dependent Magnitudes of Induced Earthquakes in the Groningen Gas Field. *Journal of Geophysical Research: Solid Earth*, *125*(11). https://doi.org/10.1029/2020jb020013

Burkhard, M., & Grünthal, G. (2009). Seismic source zone characterization for the seismic hazard assessment project PEGASOS by the Expert Group 2 (EG1b). *Swiss Journal of Geosciences*, *102*(1), 149–188. https://doi.org/10.1007/s00015-009-1307-3

Chiou, B. S.-J., & Youngs, R. R. (2008). An NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra. *Earthquake Spectra*, *24*(1), 173–215. https://doi.org/10.1193/1.2894832

Cornell, C. A., & Vanmarcke, E. H. (1969). *The major influences on seismic risk*. *1*.

Danciu, L., Giardini, D., Weatherill, G., Basili, R., Nandan, S., Rovida, A., Beauval, C., Bard, P.-Y., Pagani, M., Reyes, C. G., Sesetyan, K., Vilanova, S., Cotton, F., & Wiemer, S. (2024). The 2020 European Seismic Hazard Model: overview and results. *Natural Hazards and Earth System Sciences*, *24*(9), 3049–3073. https://doi.org/10.5194/nhess-24-3049-2024

Deichmann, N. (2017). Theoretical Basis for the Observed Break in ML/ MwScaling between Small and Large Earthquakes. *Bulletin of the Seismological Society of America*, *107*(2), 505–520. https://doi.org/10.1785/0120160318

Dyck, J. V. (1985). *Statistical Analysis of Earthquake Catalogs* [Doctoral dissertation].

Gerstenberger, M. C., Bora, S., Bradley, B. A., DiCaprio, C., Kaiser, A., Manea, E. F., Nicol, A., Rollins, C., Stirling, M. W., Thingbaijam, K. K. S., Dissen, R. J. V., Abbott, E. R., Atkinson, G. M., Chamberlain, C., Christophersen, A., Clark, K., Coffey, G. L., Torre, C. A. d. l., Ellis, S. M., … Wotherspoon, L. M. (2023). The 2022 Aotearoa New Zealand National Seismic Hazard Model: Process, Overview, and Results. *Bulletin of the Seismological Society of America*, *114*(1), 7–36. https://doi.org/10.1785/0120230182

Grünthal, G., Wahlström, R., & Stromeyer, D. (2009). The unified catalogue of earthquakes in central, northern, and northwestern Europe (CENEC)—updated and expanded to the last

millennium. *Journal of Seismology*, *13*(4), 517–541. https://doi.org/10.1007/s10950-008-9144-9

Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, *34*(4), 185–188.

Hanks, T. C., & Boore, D. M. (1984). Moment-magnitude relations in theory and practice. *Journal of Geophysical Research: Solid Earth*, *89*(B7), 6229–6235. https://doi.org/10.1029/jb089ib07p06229

Hanks, T. C., & Kanamori, H. (1979). A moment magnitude scale. *Journal of Geophysical Research*, *84*(B5), 2348. https://doi.org/10.1029/jb084ib05p02348

Johnston, A. C., Coppersmith, K. J., Kanter, L. R., & Cornell, C. A. (1994). *The earthquakes of stable continental regions* [Technical report].

Keefer, D. L., & Bodily, S. E. (1983). Three-Point Approximations for Continuous Random Variables. *Management Science*, *29*(5), 595–609. https://doi.org/10.1287/mnsc.29.5.595

Llenos, A. L., Shelly, D. R., & Shumway, A. M. (2026). Magnitude Conversion Relations Create Substantial Differences in Seismic Hazard Models. *Seismological Research Letters*. https://doi.org/10.1785/0220250231

Main, I. (1996). Statistical physics, seismogenesis, and seismic hazard. *Reviews of Geophysics*, *34*(4), 433–462. https://doi.org/10.1029/96rg02808

Main, I., Irving, D., Musson, R., & Reading, A. (1999). Constraints on the frequency–magnitude relation and maximum magnitudes in the UK from observed seismicity and glacio-isostatic recovery rates. *Geophysical Journal International*, *137*(2), 535–550. https://doi.org/10.1046/j.1365-246x.1999.00796.x

McGuire, R. K. (1995). Probabilistic seismic hazard analysis and design earthquakes: closing the loop. *Pubs.geoscienceworld.org*, *85*(5), 1275–1284.

Miller, A. C., & Rice, T. R. (1983). Discrete Approximations of Probability Distributions. *Management Science*, *29*(3), 352–362. https://doi.org/10.1287/mnsc.29.3.352

Mosca, I., Sargeant, S., Baptie, B., Musson, R. M. W., & Pharaoh, T. C. (2022). The 2020 national seismic hazard model for the United Kingdom. *Bulletin of Earthquake Engineering*, *20*(2), 633–675. https://doi.org/10.1007/s10518-021-01281-z

Musson, R. (2007). British earthquakes. *Proceedings of the Geologists' Association*, *118*(4), 305–337. https://doi.org/10.1016/s0016-7878(07)80001-0

Musson, R. M. W. (2011). *D3.7a - Assessment of activity rates for seismic source zones* (Technical Report No. Revison1; p. 23).

Musson, R. M. W. (2012). The Effect of Magnitude Uncertainty on Earthquake Activity Rates. *Bulletin of the Seismological Society of America*, *102*(6), 2771–2775. https://doi.org/10.1785/0120110224

Musson, R., & Sargeant, S. (2007). *Eurocode 8 seismic hazard zoning maps for the UK* (Technical Report No. Technical Report CR/07/125N; p. 70).

Rhoades, D. A. (1996). Estimation of the Gutenberg-Richter relation allowing for individual earthquake magnitude uncertainties. *Tectonophysics*, *258*(1–4), 71–83.

Stafford, P. J., Pettinga, J. R., & Berrill, J. B. (2008). Seismic source identification and characterisation for probabilistic seismic hazard analyses conducted in the Buller–NW Nelson Region, South Island, New Zealand. *Journal of Seismology*, *12*(4), 477–498. https://doi.org/10.1007/s10950-008-9099-x

Stromeyer, D., Grünthal, G., & Wahlström, R. (2004). Chi-square regression for seismic strength parameter relations, and their uncertainties, with applications to an M w based earthquake catalogue for central, northern and northwestern Europe. *Journal of Seismology*, *8*(1), 143–153. https://doi.org/10.1023/b:jose.0000009503.80673.51

Tinti, E., Fukuyama, E., Piatanesi, A., & Cocco, M. (2005). A Kinematic Source-Time Function Compatible with Earthquake Dynamics. *Bulletin of the Seismological Society of America*, *95*(4), 1211–1223. https://doi.org/10.1785/0120040177

Tinti, S., & Mulargia, F. (1985). Effects of magnitude uncertainties on estimating the parameters in the Gutenberg-Richter frequency-magnitude law. *Bulletin of the Seismological Society of America*, *75*(6), 1681–1697.

USNRC. (2012). *Central and Eastern United States seismic source characterisation for nulear facilities* (Technical Report No. NUREG-2115; p. 364).

Utsu, T. (1999). Representation and analysis of the earthquake size distribution: A historical review and some new approaches. *Pure and Applied Geophysics*, *155*, 509–535.

Veneziano, D., & Dyck, J. V. (1985). *Seismic Hazard Methodology for Nuclear Facilities in the Eastern U.S.* (Vol. 2, p. Appendix A – 4). EPRI/SOG.

Wason, H. R., Das, R., & Sharma, M. L. (2012). Magnitude conversion problem using general orthogonal regression. *Geophysical Journal International*, *190*(2), 1091–1096. https://doi.org/10.1111/j.1365-246x.2012.05520.x

Weichert, D. H. (1980). Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes. *Bulletin of the Seismological Society of America*, *70*(4), 1337–1346.